GENOMIC DATA-BASED MODELS OF GROWTH FACTOR SIGNALING FOR PERSONALIZED CANCER THERAPY SELECTION

by

R. Joseph Bender

A dissertation submitted to Johns Hopkins University in conformity with
the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
October 2014

Abstract

The genetic heterogeneity of cancer creates patient-to-patient variability that makes it difficult to predict whether the patient will respond to a treatment. This is particularly true for VEGF-targeting therapies, for which, in some cancers, response rates have typically been low and overall survival has only been extended slightly if at all. Development of predictive biomarkers for VEGF-targeting therapies has traditionally focused either on measurement of VEGF family ligand concentrations in the plasma or on tumor-derived transcriptomic data. Here, we incorporate both of these types of information into a whole-body computational model of the kinetic interactions between VEGF ligands and receptors. Gene expression data from a population of cancer patients allows us to create a virtual population where the effects of multiple VEGF-targeting drugs can be tested. This approach allows us to limit our analysis to relevant genes instead of the entire genome and allows us to incorporate mechanistic information, both of which should lead to models with better reproducibility.

We first examined patterns of expression of VEGF ligands and receptors as well as a related family, the Semaphorins. Several previously defined subtypes of cancer were associated with pro-angiogenic alterations in the expression of these genes. Multivariate biomarkers based on VEGF and Semaphorin gene expression were, in some cases, able to provide better separation of patients according to prognosis. These results provide clinically relevant subtypes and highlight the role that Semaphorins may play in processes that drive tumor angiogenesis and progression.

We then used gene expression data to create three virtual populations: patients with breast, kidney, or prostate cancers. Drug response metrics in these populations allowed us to determine characteristics of patients that make them more responsive to treatments. We found that the best biomarkers of response differed between cancer types. This approach can be applied to other

families of growth factors as well. Here, we demonstrate its application to the EGFR/ErbB family.

Advisor:

Feilim Mac Gabhann, Assistant Professor, Department of Biomedical Engineering & Institute for Computational Medicine, The Johns Hopkins University

Thesis Committee (alphabetically):

Donald Geman, Professor, Department of Applied Mathematics and Statistics & Institute for Computational Medicine, The Johns Hopkins University

Hans Hammers, Assistant Professor, Department of Oncology, The Johns Hopkins University

Acknowledgments

I would like to acknowledge my advisor, Feilim Mac Gabhann, for his guidance during my time working in his lab. Under his mentorship I have grown tremendously as a scientist. He has imparted immeasurable skills beyond scientific research that I will take with me in my career, and I have really enjoyed being a part of his lab these past four years.

I'd also like to thank the members of my thesis committee, Dr. Don Geman and Dr. Hans Hammers, for their guidance and feedback.

To Iraj, Lindsay, Liz, Laura, and the other members of the Mac Gabhann lab, thank you for your support and friendship. In particular, I would like to recognize Sean and Adriana, two undergraduates whom I was very fortunate to work with during my time at Hopkins.

Additionally, thanks to my parents, Bob and Regina, for their love and support throughout my life, and particularly in my academic pursuits.

Finally, I dedicate my thesis to my wife, Rachael. Without her, I would not be where I am today. I can't thank you enough for all the support you've given me.

The content of Chapter 2 was previously published as “Expression of VEGF and Semaphorin genes defines subgroups of triple negative breast cancer” in *PLoS One.*

The content of Chapter 4 is currently in review.

Manuscripts for the content in Chapters 5 and 7 are currently in preparation.

Table of Contents

List of Figures

List of Tables

1 Introduction

The genomic heterogeneity of cancer produces widely varying responses to therapies, not only between different types of cancer, but also from patient to patient with the same type of cancer. Therapies that target angiogenesis, the growth of new blood vessels from existing vasculature, in particular suffer from low response rates. Biomarkers are needed that can be used at the time of diagnosis to differentiate tumors likely to respond to a drug tumors likely to be resistant. At least two key types of information can aid in the development of predictive biomarkers: 1) molecular-level mechanistic knowledge of the pathway targeted by a therapy, and 2) measurements of the level of expression of genes in that pathway, from a large set of tumor samples. Currently only the latter is typically used; we will integrate both together to improve the predictive power.

1.1 Tumor Angiogenesis

In order to support the rapidly expanding population of tumor cells that is characteristic of cancer, an adequate supply of oxygen and other nutrients must be maintained. To do this, tumors secrete proteins into the microenvironment that promote angiogenesis, the formation of new blood vessels from existing networks of capillaries. This blood vessel sprouting and remodeling, which is a normal part of organ growth and of adult physiology, can be co-opted to supply tumors by stimulating the growth of new branches from the host organ vasculature. Due to its importance in tumor progression, angiogenesis is considered a hallmark of cancer [1].

Angiogenesis is a complex multicellular process. Endothelial cells are the major cell type comprising blood vessels. Sprout formation from existing blood vessels occurs due to proliferation and migration of endothelial cells. Chemical signals originating from tumor cells create gradients that guide the direction of endothelial cell migration. Other stromal cells in the tumor microenvironment contribute to angiogenesis as well. Pericytes associate with and stabilize

microvessels composed of endothelial cells. Cells from the immune system, such as macrophages, are recruited to the site of the tumor and secrete pro-angiogenic chemical signals [2,3].

The wide array of chemical signals can influence angiogenesis, either by enhancing or inhibiting it. The shift in the net effect of these factors from anti-angiogenic to pro-angiogenic that occurs during tumor progression is referred to as the angiogenic switch [4]. A major pro-angiogenic group of proteins is the vascular endothelial growth factor (VEGF) family (Section 1.1.2). Other pro-angiogenic growth factors include the fibroblast growth factor (FGF) family, the platelet-derived growth factor (PDGF) family, the hepatocyte growth factor (HGF) family, the epidermal growth factor (EGF) family, and the angiopoietin family. Inhibitors of angiogenesis include thrombospondins [5] and peptides derived from CXC chemokines and type IV collagen [6]. Another family, the semaphorins (Section 1.1.3), have both pro- and anti-angiogenic effects, depending on the specific member of the family.

1.2 Molecular Regulators of Angiogenesis in Cancer

The complexity of the molecular regulatory network for angiogenesis – the number of competing ligands and receptors, the synergistic and antagonistic downstream effects – is exacerbated by the observation that all of the components can be dysregulated in cancer. It would be impossible to integrate all these changes and the underlying interaction network to predict outcomes without a detailed mathematical model.

1.2.1 VEGF Family

The VEGF family plays a large role in the regulation of angiogenesis. This family comprises five genes encoding ligands (VEGFA, VEGFB, VEGFC, VEGFD, and PlGF), three receptor tyrosine kinases (VEGFR1, VEGFR2, and VEGFR3), and two neuropilin co-receptors (NRP1 and NRP2). The network of interactions between these families is complex. Some of the ligand genes encode multiple splice isoforms, with the $VEGFA_{121}$, $VEGFA_{165}$, $VEGFA_{189}$, PlGF-

1, PlGF-2, $VEGFB_{167}$, and $VEGFB_{186}$ proteins playing major roles in adult angiogenesis. Proteolytic processing of the ligands plays a role in modulating the receptor binding affinity of VEGFC and VEGFD. Alternative splicing also affects the genes encoding receptors: shortened soluble forms of VEGFR1 and VEGFR2 that retain ligand binding have been found. Some ligand isoforms are heparin-binding and thus reversibly bind to extracellular matrix, from which they can be released by proteases. The isoforms also have differing affinities for neuropilins, with some unable to bind them at all. Homodimers of all three RTKs occur, as well as VEGFR1-VEGFR2 and VEGF2-VEGFR3 heterodimers. Complex formation between VEGF RTKs and neuropilins can occur in a ligand-dependent or -independent manner, depending on the ligand and receptor. The VEGF receptors can form complexes with neuropilin co-receptors, altering both ligand affinities and signaling output [7,8].

Ligand-activated VEGF receptors induce a cascade of intracellular signals that ultimately promote the proliferation and migration of endothelial cells. VEGF receptors are type V RTKs with seven extracellular immunoglobulin domains, a transmembrane domain, and intracellular tyrosine kinase domains. Ligand binding triggers receptor dimerization, which leads to receptor phosphorylation, recruitment of adapter proteins, and downstream signaling. Phosphorylation of specific tyrosine residues on the receptors leads to differences in phenotype due to activation of different pathways, with phosphorylation of VEGFR2 at Y951, Y1175, and Y1214 in particular having varying abilities to lead to proliferation, migration, survival, and permeability. The role of VEGFR1 phosphorylation and signaling is unclear; some studies indicate VEGFR1 may inhibit angiogenesis simply by serving as a VEGF sink, leading to reduced VEGFR2 binding. VEGFR2 signaling tends to promote angiogenesis while VEGFR3 tends to promote lymphangiogenesis due to its expression on lymphatic endothelial cells [7,8].

1.2.2 Semaphorin Family

Semaphorins are a large family of 20 proteins with diverse roles in both neural and vascular development. They are organized into 5 subfamilies (classes 3 through 7) based on their structures. The class 3 semaphorins are unique in two ways: they are the only subfamily that can bind to the VEGF co-receptors neuropilin-1 and -2, and they are secreted in a soluble form. The other subfamilies are membrane-associated, although proteolytic cleavage can produce soluble forms. The receptors responsible for semaphorin signal transduction are the plexins. Whereas class 3 semaphorins must bind to a neuropilin co-receptor prior to forming a complex with plexins, semaphorins in the other classes bind plexins directly. Plexins themselves form a large family of proteins; there are 9 plexin receptors, divided into 4 subfamilies (classes A through D). Plexins have intracellular GAP domains that can inactivate R-Ras, leading to collapse of the cytoskeleton and reduced motility [9].

Semaphorins were originally studied to gain insight into their functions in neural development but more recently their effects on the tumor vasculature have been characterized [9]. Knockdown of Sema3A in mouse models increases angiogenesis, whereas addition of exogenous Sema3A reduced the vascular density [10]. Most class 3 semaphorins have the ability to repel endothelial cells in culture [11], with some dependence on the specific neuropilin and plexin receptors that are present [12]. Semaphorins outside of class 3 also play roles in angiogenesis. Sema4A inhibits angiogenesis in chick embryos and mouse corneas by binding to Plexin-D1 and inhibiting the formation of Rac-GTP [13]. Some semaphorins are able to enhance angiogenesis. Sema4D can increase endothelial cell motility by binding to Plexin-B1 and activating a variety of kinases including Src [14]. Sema5A can enhance angiogenesis through binding to Plexin-B3 by activating endothelial cell proliferation and motility [15].

A potential mechanism of anti-angiogenesis for class 3 semaphorins is through competitive displacement of VEGF from neuropilin co-receptors. Competitive displacement

between $VEGF_{165}$ and Sema3A has been demonstrated in neuropilin-1-overexpressing COS-7 cells [16], but this conflicts with data suggesting that the binding sites of these two ligands on neuropilin-1 are distinct [17]. More recent data indicates that proteolytic processing of class 3 semaphorins controls their ability to competitively displace VEGF by exposing a C-terminal arginine residue that interacts with the VEGF-binding site on neuropilin-1 [18,19].

1.3 Therapeutic Targeting of Angiogenesis in Cancer

Due to its importance in tumor angiogenesis, the VEGF pathway has been the target of multiple cancer therapies. Monoclonal antibodies have been used to target the pathway outside of cells, including the anti-VEGF antibody bevacizumab, the anti-VEGFR2 antibody ramucirumab, and the anti-NRP1 antibody MNRP1685A. Inside cells, small molecule tyrosine kinase inhibitors (TKIs) have been used to prevent VEGF receptor phosphorylation. FDA-approved VEGF receptor-targeting TKIs include sorafenib, sunitinib, pazopanib, and cabozantinib.

VEGF-targeting therapies have had varying levels of success in clinical trials. The VEGF-neutralizing antibody bevacizumab is currently approved for treatment of colorectal [20], lung [21], brain [22,23], kidney [24,25], and cervical [26] cancers based on its ability to extend overall survival. Bevacizumab has failed to extend overall survival in pancreatic [27], breast [28,29], ovarian [30], and prostate [31] cancers. Despite extending the overall survival, overall response rates to bevacizumab are fairly low even in successful clinical trials, with 31% to 48% of patients responding. Although this represents a 10% to 20% improvement over placebo-treated patient cohorts, the low response rates demonstrate the need for improved patient selection. Similarly low response rates have been observed for ramucirumab in gastric cancer [32] and for the various TKIs in kidney [33], pancreatic [34], gastrointestinal stromal [35], and liver [36] cancers. A goal of personalized medicine is to create predictive biomarkers that would allow patients to be grouped into likely responders and non-responders at diagnosis; this would drastically improve response rates.

Table 1-1: Bevacizumab clinical trial results.

Clinical trial data for response rates (RR, includes complete and partial responses), progression-free survival (PFS), and overall survival (OS).

Indication	*N*	RR	Median PFS (months)	Median OS (months)	Ref
Colorectal	813	44.8% vs. 34.8% ($P = 0.004$)	10.6 vs. 6.2 ($P < 0.001$)	20.3 vs. 15.6 ($P < 0.001$)	[20]
NSCLC	878	35% vs. 15% ($P < 0.001$)	6.2 vs. 4.5 ($P < 0.001$)	12.3 vs. 10.3 ($P = 0.003$)	[21]
Glioblastoma[1]	85	28.2%	4.2	9.2	[23]
Renal Cell Carcinoma	649	31% vs. 13% ($P = 0.0001$)	10.2 vs. 5.4 ($P = 0.0001$)	23.3 vs. 21.3 ($P = 0.336$)	[24, 25]
Cervical	452	48% vs. 36% ($P = 0.008$)	8.2 vs. 5.9 ($P = 0.002$)	17.0 vs. 13.3 ($P = 0.004$)	[26]
Gastric	774	46.0% vs. 37.4% ($P = 0.0315$)	6.7 vs. 5.3 ($P = 0.0037$)	12.1 vs. 10.1 ($P = 0.10$)	[37]
Pancreatic	535	13% vs. 10% *N.S.*	3.8 vs. 2.9 ($P = 0.07$)	5.8 vs. 5.9 ($P = 0.95$)	[27]
Ovarian	1,528	67% vs. 48% ($P < 0.001$)	24.1 vs. 22.4 ($P = 0.04$)	NA	[30]
Breast	722	36.9% vs. 21.2% ($P < 0.001$)	11.8 vs. 5.9 ($P < 0.001$)	26.7 vs. 25.2 ($P = 0.16$)	[29]
Triple-Negative Breast (adjuvant)[2]	2,591	NA	83.7% vs. 82.7% (3-year PFS, $P = 0.18$)	93% vs. 92% (at data cutoff, $P = 0.23$)	[38]
Prostate	1,050	49.4% vs. 35.5% ($P = 0.0013$)	9.9 vs. 7.5 ($P < 0.001$)	22.6 vs. 21.5 ($P = 0.181$)	[31]

[1] Single-arm trial

[2] Median PFS and OS not reached because this was an adjuvant trial in non-metastatic disease, therefore PFS and OS rates are listed.

No biomarkers have been found that are able to completely distinguish responders from non-responders for these drugs. Patients with higher baseline levels of plasma VEGF sometimes have better response rates to bevacizumab, but this trend does not always hold [39]. High baseline plasma sVEGFR2 [40] and low tumor neuropilin-1 [41] have also had associations with improved bevacizumab responsiveness in some trials. The failure of a single metric or univariate biomarker to adequately distinguish patients may be the result of the complexity of the VEGF pathway and other regulators of angiogenesis. Multivariate biomarkers that account for the levels of multiple

components of the VEGF network may provide a more robust solution to the problem of predicting responders.

1.4 Mathematical Models of Growth Factor Ligand-Receptor Networks

Mathematical modeling has been used to further our understanding of the complexities of the VEGF pathway. Early models described the binding of $VEGFA_{165}$ and $VEGFA_{121}$ to the receptors VEGFR1, VEGFR2, and NRP1. The models comprise a system of nonlinear differential equations that included experimentally measured kinetic rate constants for ligand binding and receptor coupling reactions. This modeling approach has allowed for prediction of the effects of competitive PlGF binding on VEGF binding distributions [42], and has also led to an understanding of the mechanism by which NRP1 enhances the binding of $VEGFA_{165}$ to VEGFR2 [43].

Computational models of the VEGF pathway have been expanded to include multiple cell types in multiple tissues of the body. These compartment models typically include three compartments: a diseased tissue, blood, and normal tissue. The normal compartment is a generic compartment comprising all normal tissues in the body, and the blood allows for transport between normal and diseased tissues. Reactions describing the transport of the soluble proteins in the model between compartments have been added. Diseased tissues have included tumors and ischemic muscles [44]. The role of a soluble form of VEGFR1, sVEGFR1, has been investigated in the context of compartment models [45]. Recently, the anti-VEGF antibody bevacizumab has been added into the model to create whole-body pharmacokinetic models [46,47]. In this work, we leverage models such as these. We expand the molecular complexity, and go beyond an 'average patient' approach to create populations of individualized models. These will be useful in developing predictive biomarkers of VEGF-targeting therapies. They provide a framework in which the effects of inter-patient variability on drug response can be tested.

1.5 Cancer Bioinformatics

Microarray and sequencing technologies have yielded insight into the abundances of molecules on a genome-wide scale. Efforts such as the Cancer Genome Atlas (TCGA) have produced large datasets from multiple cancer types describing gene copy number, expression, and methylation, as well as microRNA expression, protein expression, and mutations.

1.5.1 Molecular Subtypes of Cancers

Genomic data has been used to define molecular subtypes for many cancer types. In breast cancer, for example, hierarchical clustering was used to define the basal, luminal A, luminal B, and HER2-enriched molecular subtypes based on the expression of an "intrinsic" subset of genes that displayed high variability between tumor samples but low variability between different samples taken from identical tumors [48,49]. A classifier based on nearest centroid classification of 50 genes was developed, giving these subtypes the name "PAM50" [50]. Recently, an integrative analysis of copy number, DNA methylation, mRNA, microRNA, and protein measurements from the TCGA dataset yielded clusters that significantly overlapped with the PAM50 subtypes [51], reinforcing their utility. These molecular subtypes also overlap significantly with the level of three cell surface receptors that are evaluated using immunohistochemistry: estrogen receptor (ER), progesterone receptor (PR), and HER2. Tumors that are negative for all three of these receptors ("triple-negative") tend to have poorer prognoses due to a more invasive phenotype and fewer treatment options [52]. Triple-negative breast cancers (TNBC) tend to fall into the basal subtype, while ER+ tumors tend to be either luminal A or B. The PAM50 subtypes have prognostic significance and limited significance as predictors of treatment outcome.

Molecular subtypes have been defined in other cancer types based on TCGA data. Hierarchical clustering was used to divide glioblastoma samples into four subtypes [53,54] and to divide colorectal cancers into three subtypes [54]. Non-negative matrix factorization was used to

find four subtypes of ovarian cancer [55]. The molecular subtypes found in these studies all separated patients into groups with significantly different survival times. Beyond prognosis, however, further studies are needed to determine what role, if any, the subtypes may have in therapy selection.

1.5.2 Predictive Biomarkers for Treatment

Predictive biomarkers for targeted therapies have generally been unsuccessful. Notable exceptions include HER2 status, which is predictive of response to HER2-targeting drugs such as trastuzumab [56], pertuzumab [57], and lapatinib [58]. Tyrosine kinase inhibitors are sometimes effective only in tumors harboring a particular mutation, such as BRAF V600E for vemurafenib [59]. Multivariate gene expression signatures are not yet used for molecularly targeted therapies, but some have had success in predicting response to chemotherapy. The 21-gene OncotypeDx test for breast cancer is predictive of benefit from chemotherapy in ER+ patients [60].

There are several pitfalls in developing predictive biomarkers from genome-wide expression data. Simple linear, univariate biomarkers based on the level of expression of the target molecule often fail to adequately capture the complexity of signaling pathways. On the other hand, biomarkers learned from datasets containing measurements of thousands of genes may suffer from overfitting. Knowledge-based biomarkers may solve many problems associated with biomarker development. Extensive knowledge of the connections between molecules and the dynamics of their interactions derived from experiments can be used to inform the biomarker development process. The set of potential predictor variables can be limited to molecules relevant to the pathway of interest. Equations describing the rates of interactions between molecules in the pathway can be used to introduce biologically relevant nonlinearities into biomarkers as well.

2 Expression of VEGF and Semaphorin Genes Defines Subgroups of Triple Negative Breast Cancer

2.1 Summary

Triple negative breast cancers (TNBC) are difficult to treat due to a lack of targets and heterogeneity. Inhibition of angiogenesis is a promising therapeutic strategy, but has had limited effectiveness so far in breast cancer. To quantify heterogeneity in angiogenesis-related gene expression in breast cancer, we focused on two families – VEGFs and semaphorins – that compete for neuropilin co-receptors on endothelial cells. We compiled microarray data for over 2,600 patient tumor samples and analyzed the expression of VEGF- and semaphorin-related ligands and receptors. We used principal component analysis to identify patterns of gene expression, and clustering to group samples according to these patterns. We used available survival data to determine whether these clusters had prognostic as well as therapeutic relevance. TNBC was highly associated with dysregulation of VEGF- and semaphorin-related genes; in particular, it appeared that expression of both VEGF and semaphorin genes were altered in a pro-angiogenesis direction. A pattern of high VEGFA expression with low expression of secreted semaphorins was associated with 60% of triple-negative breast tumors. While all TNBC groups demonstrated poor prognosis, this signature also correlated with lower 5-year survival rates in non-TNBC samples. A second TNBC pattern, including high VEGFC expression, was also identified. These pro-angiogenesis signatures may identify cancers that are more susceptible to VEGF inhibition.

2.2 Methods

2.2.1 Data sets

Published human breast cancer gene expression data sets were collated based on the following criteria: the tumors had to be untreated, primary tumors, and the gene expression had to be analyzed using the Affymetrix GeneChip® Human Genome U133A platform. Of the 98 data sets returned by searching for human breast tumors on the U133A platform in the GEO database, 22 met the criteria of being untreated and primary as of April 13, 2012. If available, the following data were also collected: ER, PR, and HER2 immunohistochemistry (IHC), lymph node status, age at diagnosis, tumor stage, and tumor grade. The breast cancer data sets [61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80] (**Table 2-1**) were compiled into one expression data set and normalized using the *justRMA* function in the *affy Bioconductor* package of the R statistical software environment. Some samples were removed prior to normalization: 30 samples in GSE20194 were replicates, 47 samples in GSE5847 were stromal cells isolated by laser capture microdissection, and 20 samples in GSE5847 had received neoadjuvant chemotherapy prior to surgery when the sample was taken. Many of the samples in different data sets were found to be from the same patients; samples were removed so that each patient was represented in the final data set only once. After removal of samples, the data set consisted of 2,656 individual tumor samples and 42 normal samples. When multiple probe sets corresponded to a single gene, only the probe set with the highest variance across all samples was used to represent expression of the gene. In total, we analyzed the expression from 55 probe sets corresponding to 33 genes (**Figure 2-1**).

The two TCGA data sets used for validation were current as of April 25, 2012. One data set consisted of 537 tumor samples analyzed on the Agilent G4502A microarray platform, while the other was an RNA-Seq data set consisting of 750 tumor samples analyzed on the Illumina HiSeq 2000 system. There were 481 patients overlapping between these two TCGA data sets.

This replication was allowed in order to show similar results using different gene expression measurement technologies. Positive and negative status for the receptors was assigned based on IHC if available, otherwise based on the gene expression measurements for the ESR1, PGR, and ERBB2 genes.

Table 2-1: Breast cancer gene expression datasets

As noted in the main text, samples must be untreated primary tumors. Unless otherwise noted, each sample in the datasets represents one tumor. The numbers of tumor samples are the actual number of samples used in the analysis; replicate samples were removed.

Dataset	N	Reference	Notes
GSE1456	159	65	
GSE1561	49	66	Core biopsy, >20% tumor cell content
GSE2034	286	67	Tumor cell content >70%, all lymph-node negative
GSE2603	99	68	Tumor cell content >70%
GSE2990	104	69	
GSE3494	251	70	
GSE5327	58	71	All ER-
GSE5847	28	72	47 stroma, 48 tumor (LCM); Surgical samples
GSE7390	198	73	
GSE11121	200	74	Tumor cell content >40%
GSE20194	42	75	Tumor cell content >70%, 30 replicates
GSE20271	116	76	
GSE20437	42	77	Normal breast tissue (no tumors)
GSE21217	11	78	Surgical samples
GSE22093	68	79	
GSE22597	74	53	
GSE23988	61	79	
GSE24185	103	80	
GSE25066	508	81	
GSE31519	67	82	
GSE32072	25	83	
GSE36772	100	N/A	
GSE36773	49	N/A	

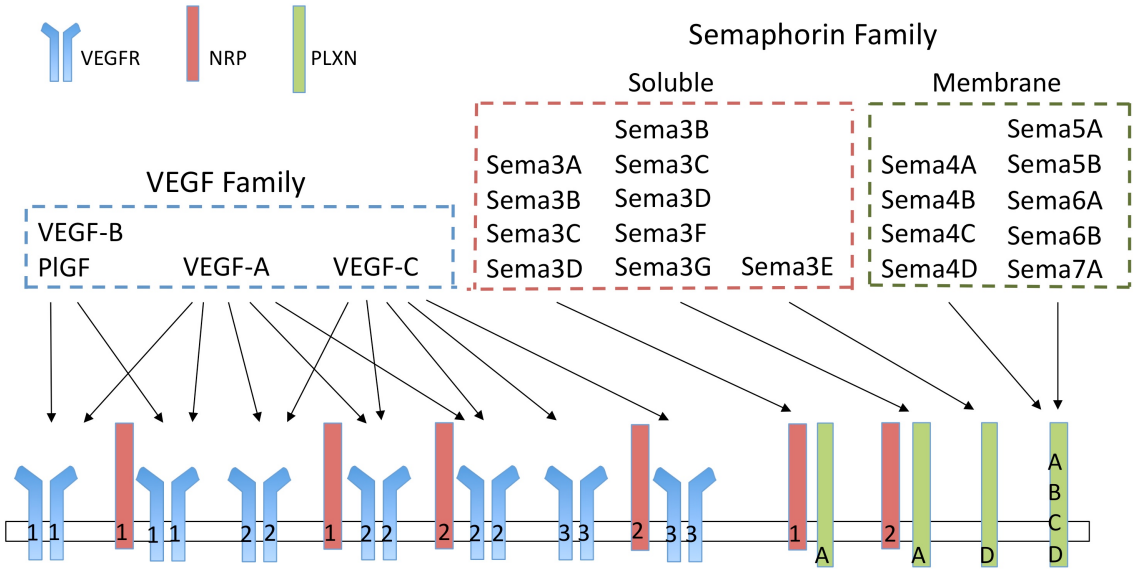

Figure 2-1: Ligand-Receptor interactions for the VEGF and Semaphorin families.

VEGF ligands bind to and signal through three RTKs: VEGFR1, VEGFR2, and VEGFR3 (blue). Neuropilins are in red, with numbers to distinguish between neuropilin-1 and neuropilin-2. Semaphorin ligands bind to and signal through Plexins A-D (green). Many (but not all) members of the VEGF and Sema3 families use Neuropilin 1 or 2 as a co-receptor for binding to the canonical signaling receptors. This competition for Neuropilin is thought to represent one mechanism by which VEGF and Semaphorin ligands antagonize each other; in addition, the downstream signaling of VEGFRs and Plexins can have opposite function. Note that not all splice isoforms of VEGF-A, VEGF-B, and PlGF can bind to the receptors indicated, and that kinetic rates of binding vary among isoforms.

2.2.2 Assignment of TN status and TN subtypes

The assignment of triple negative status and subtype was made based on the gene expression levels of ESR1, PGR, and ERBB2 when IHC data were not available, as previously described by others [81]. A Gaussian distribution was fit based on expression of the three receptors for IHC TN samples, and another distribution was fit for IHC receptor-positive samples. Samples with no IHC data were classified by computing the probability of being TN based on the two density functions derived from samples with IHC data. Comparison of IHC data to expression-based assignments has demonstrated that misclassification of samples is rare (<3.6%) [81].

Subtypes of the triple negative classification were assigned by calculating subtype centroids based on the classification used previously by others [50]: expression of ~2000 genes was used to compute centroids of each of the six subtypes based on 193 tumor samples. The classifier derived from this was tested using leave-one-out cross validation and classified 171 of the 193 samples correctly, for an accuracy of 88.6%. Testing the classifier trained with all 193 samples resulted in correct classification of 187 samples, for an accuracy of 96.9%. This classifier was used to determine the subtypes of the remaining 582 triple negative samples.

2.2.3 PAM50 intrinsic subtypes

A previously used classifier for breast cancer involves the use of 50 genes to place tumors into one of five categories: basal, luminal A, luminal B, HER2-like, and normal-like. The method for classifying a new sample is to take the Spearman correlation coefficient of the expression of the 50 genes in the sample with each of the five class centroids. The class whose correlation coefficient is the highest is the class to which the sample belongs, unless all correlation coefficients are less than 0.1, in which case the sample is unclassified [82]. It should be noted that no genes in the PAM50 classifier overlap with the VEGF- and semaphorin-related

genes that we consider here; thus when we compare VEGF- and semaphorin-based clusters, we are considering two completely independent methods of classification.

2.2.4 Differential expression

Genes for VEGF and semaphorin ligands and receptors that were significantly different between two groups (e.g. tumor vs. normal, receptor-positive vs. triple negative, etc.) were determined by the Wilcoxon rank sum test. This was carried out using the *wilcox.test* function in R.

2.2.5 Principal component analysis

Principal component analysis (PCA) was used to reduce the dimensionality of the data sets from the 31 VEGF- and semaphorin-related genes under consideration to a smaller number of components that can reproduce most of the variability in the data. The components are linear combinations of the expression of the genes, and capture patterns of co-expression. The *prcomp* function in R was used to perform PCA. The columns of the *x* matrix returned by this function corresponded to the scores, while the columns of the *rotation* matrix corresponded to the gene loadings. For 2-D score plots where colors were used to show different groups of samples, the statistical significance of differences in PCA scores between the groups was determined using multivariate analysis of variance. The p-values were determined by comparing the Wilk's lambda statistic to a chi-squared distribution.

2.2.6 Logistic regression

For triple-negative status, lymph node status, tumor stage, tumor grade, age at diagnosis, and tumor size, logistic regression models were fit based on the scores of the first eight principal components using the R function *glm*.

2.2.7 Survival analysis

The R package *survival* was used to perform survival analysis on tumor samples for which survival and clinical variables were available. A log-rank test was used to assess univariate

significance of factors. A Cox proportional hazards model was used for multivariate analysis of all factors.

2.2.8 Cluster analysis

K-means clustering was performed on the 2656-sample data set consisting of all of the breast tumors, as well as the 775-sample data set consisting of all triple negative tumors. The R function *kmeans* was used for clustering. To ensure that the algorithm converged to the global minimum instead of a local minimum, clustering was performed 50 times and the solution with the lowest within-class sum of squares was used to determine the cluster membership of each sample.

Consensus *K*-means clustering was used to assess the stability of the clusters. This consisted of performing the clustering algorithm 100 times on different subsets of the data set, and then computing the fraction of iterations in which any pair of samples were found in the same cluster. At each iteration, the sample subsets were determined by taking a random sample without replacement whose size was 80% of the data set. The consensus matrix is a visual representation of the fraction of iterations in which any pair of samples co-clustered. The cumulative distribution of the consensus matrix across all possible sample pairs was used to determine the number of clusters. The appropriate number of clusters was the cluster number at which no further increases in the area under the cumulative distribution curve occurred. Typically, the relative change in area is close to zero above a certain value of K. For example, in the all-tumor data set and the TNBC-only data set, values of K greater than and equal to 5 resulted in low relative area changes. To select the appropriate number of clusters from these cases, the consensus matrices were investigated to determine which cluster number resulted in the most off-diagonal white space [82].

2.2.9 Visualization of data

Gene expression differences between tumor and normal samples were plotted in **Figures 2-2A** and **2-2B** as the log of the ratio of the two means. Error bars corresponded to the 99% confidence interval of the log ratio derived from the Wilcoxon rank sum test. The range of gene expression across groups was shown in boxplots with the extreme ends of the boxes corresponding to the 25^{th} and 75^{th} percentile of the data and the line inside the box corresponding to median. The whiskers extended to the furthest point outside of the boxes that still fell within 1.5 times the interquartile range from the nearest end of the box, where the interquartile range was the difference between the 75^{th} and 25^{th} percentiles.

Heatmaps of gene expression data were generated in R using the *image* function. Data were scaled (zero-mean, unit-variance) and assigned colors, with red corresponding to high expression and green corresponding to low expression. Ordering of genes in the heatmaps was performed using the *hclust* function in R with the complete-linkage agglomeration method. When dendrograms were used, they were generated using the *plot.dendrogram* function in R.

2.3 Results

2.3.1 Expression patterns defining tumor and normal tissue

The 2,656 tumor samples were compared to 42 normal samples to identify differentially expressed genes. Overall, 20 out of 29 of the ligand probe sets differed between tumor and normal tissues at a significance level of 0.01 (**Figure 2-2A**). All of the VEGF ligands were up-regulated in tumors except PlGF. PlGF had two probe sets on the U133A platform: one was down-regulated and the other did not differ significantly between normal and tumor tissues. Semaphorin ligand expression differences varied, with a mix of up- and down-regulation. Ligands in Figure 2-2A were annotated on the left axis with a color of red if they were known to have pro-angiogenic functions, blue for anti-angiogenic functions, gray if they had been shown to both promote and inhibit angiogenesis in different studies, or white if no data were available

[8,11,12,13,14,15,83,84,85,86,87,88,89,90,91,92,93,94,95,96]. In general, class-3 semaphorins inhibit angiogenesis, while the other four classes have varying effects.

Although 20 out of the 26 receptor probe sets differed from normal expression at a significance level of 0.01 (**Figure 2-2B**), the overall magnitude of the fold change between tumor and normal tissue was less for receptors than for ligands: the mean absolute difference was 0.58 for ligands (mean of all points in Figure 2-2A) compared to 0.32 for receptors (mean of all points in **Figure 2-2B**). Most receptors that were differentially expressed were up-regulated, with the exception of FLT4 (VEGFR3) and PLXNA1.

We used principal component analysis (PCA) to determine patterns of covariation in gene expression between tumor and normal samples. Projection of the gene expression data onto the plane defined by principal components 1 and 4 (PC1, PC4) showed the best separation between tumor and normal samples (**Figure 2-2C**). Normal samples showed moderate values of PC1 and high values of PC4; tumor samples had a broader range of PC1 values and lower values of PC4. High PC1 scores were associated with high expression of VEGFC, KDR, NRP1, and PLXNC1, and low expression of SEMA3A, SEMA7A, FLT1, and FLT4. Low PC1 scores were associated with the opposite expression pattern. Low PC4 scores were associated with a pro-angiogenesis signature, a combination of high expression of VEGFA, SEMA4D, NRP2, and PLXNA1, and low expression of several secreted semaphorins: SEMA3B, SEMA3C, SEMA3E, SEMA3F, and SEMA3G. The advantage of the PCA-based approach for comparing tumor and normal samples over the differential expression analysis used in Figures 2-2A and 2-2B was that patterns of co-expression could be observed. For example, KDR expression by itself was not significantly altered between tumor and normal samples (**Figure 2-2B**), but co-expression of KDR with VEGFC, NRP1, and PLXNC1 was associated with tumors (**Figure 2-2D**).

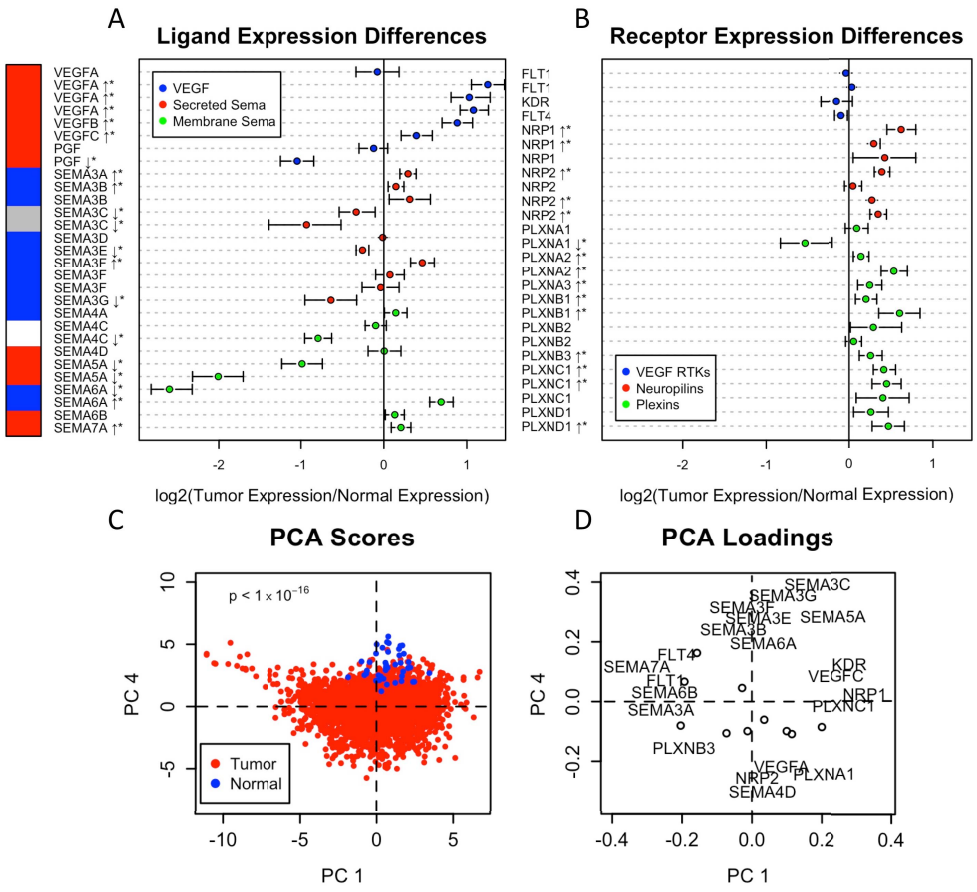

Figure 2-2: Differences in expression patterns of VEGF- and semaphorin-related genes between normal breast tissue (n=42) and breast tumors (n=2,656).

A-B, Differences in mean expression with 99% confidence intervals as determined by the Wilcoxon rank sum test for (A) ligands and (B) receptors at the probe level. Ligands are marked with the following colors to denote known effect on angiogenesis: red for pro-angiogenic, blue for anti-angiogenic, gray for context-dependent (could be pro-or anti-angiogenic) and white for unknown. Genes for which expression is significantly altered in tumors are denoted by * ($p<0.001$) and the direction noted by an arrow. **C-D**, Principal component analysis shows separation of tumors from normal samples based on first and second principal component scores (C), with corresponding gene expression patterns given by the loadings for these components (D). Gene labels are only shown for genes whose loading vectors onto PC1 and PC4 exceed a magnitude of 0.23. Circles denote the loading of genes whose names do not appear.

2.3.2 VEGF and semaphorin gene expression are differentially regulated in triple negative breast cancer

To determine patterns in VEGF and semaphorin expression that may be important in distinguishing various breast cancer subgroups, we performed PCA on the expression measurements for the 31 VEGF- and semaphorin-related genes in the data set consisting of 2,656 tumors. We compared the scores obtained from PCA with commonly used clinical variables and found that the principal components had the most significant associations with triple negative status (**Figure 2-3C**), as indicated by the large logistic regression coefficients. Some significant associations were found between the principal components and lymph node status (**Figure 2-3D**) and tumor grade (**Figure 2-3E**), but the coefficients were much smaller than those for triple negative status. Tumor stage was not associated with the components at all (**Figure 2-3F**). Additionally, we noted that applying PCA to the VEGF-related gene subset alone failed to distinguish TNBC samples from receptor-positive samples as effectively as the combined data set (**Figure 2-3A**). Applying PCA to the semaphorin-related gene subset alone resulted in some significant associations with triple-negative status (**Figure 2-3B**). The NRP1 and NRP2 genes were included in both subsets. Together, this suggests that the (indirect) interactions between the VEGFs and the semaphorins lead to different neuropilin-regulated signaling activities in TNBC tumors compared to receptor-positive tumors.

When applying PCA to the combined VEGF and semaphorin data set, the projection of the data onto the fourth principal component (PC4a, "a" to denote the all-tumor data set) provided the highest degree of separation between TNBC samples and the rest of the tumors (**Figure 2-4A**), with low values of PC4a corresponding to TNBC samples. A group of tumors also scored highly on PC3a. A relatively large proportion of these were TNBC samples that did not score low on PC4a. PC1a had a slight association with TNBC status but only in samples that scored low on PC4a, while PC2a did not appear to have any association with TNBC status. PC4a also was

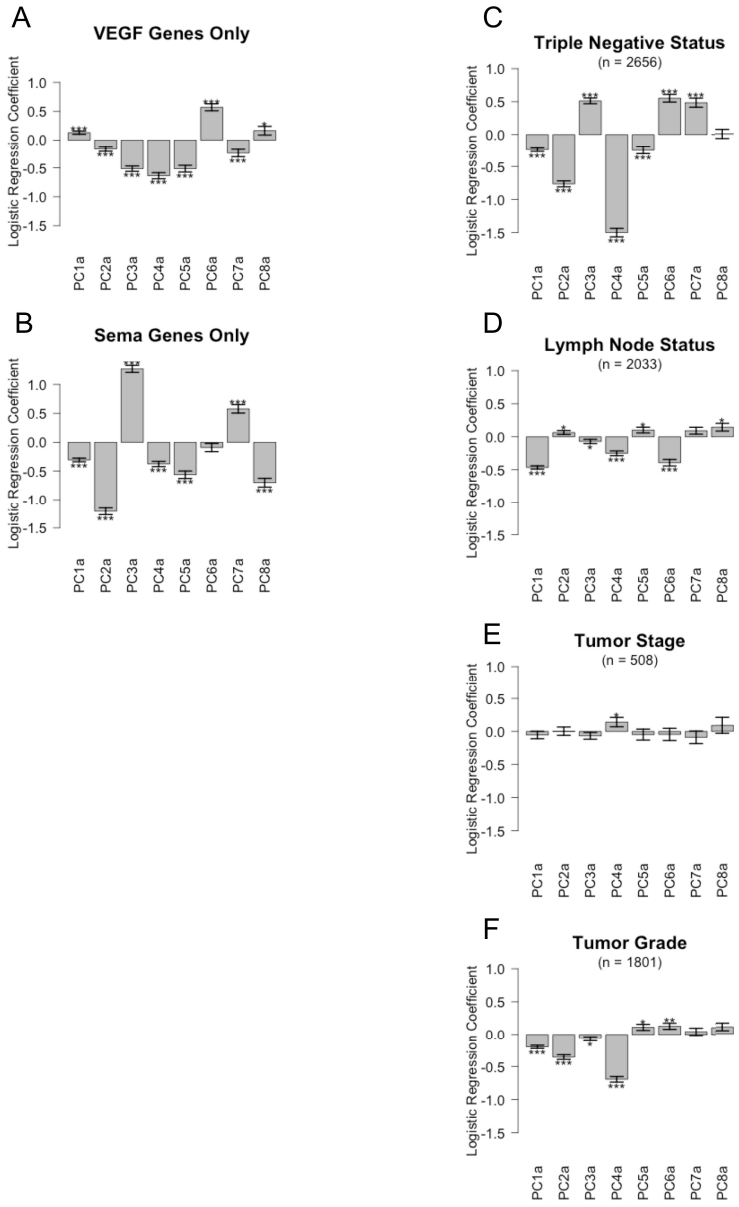

Figure 2-3: Relationship between PCA scores and TN status.

A-B, Logistic regression coefficients for the first eight PCA scores for VEGF-related genes only (A) and for the semaphorin related genes only (B). The probe sets for NRP1 and NRP2 were included in both subsets of the data. **C-F,** Logistic regression coefficients for the combined VEGF/Sema set for triple negative status (C), lymph node status (D), tumor stage (E), and tumor grade (F). The value of n in C-F indicates how many samples had the relevant annotated data available.

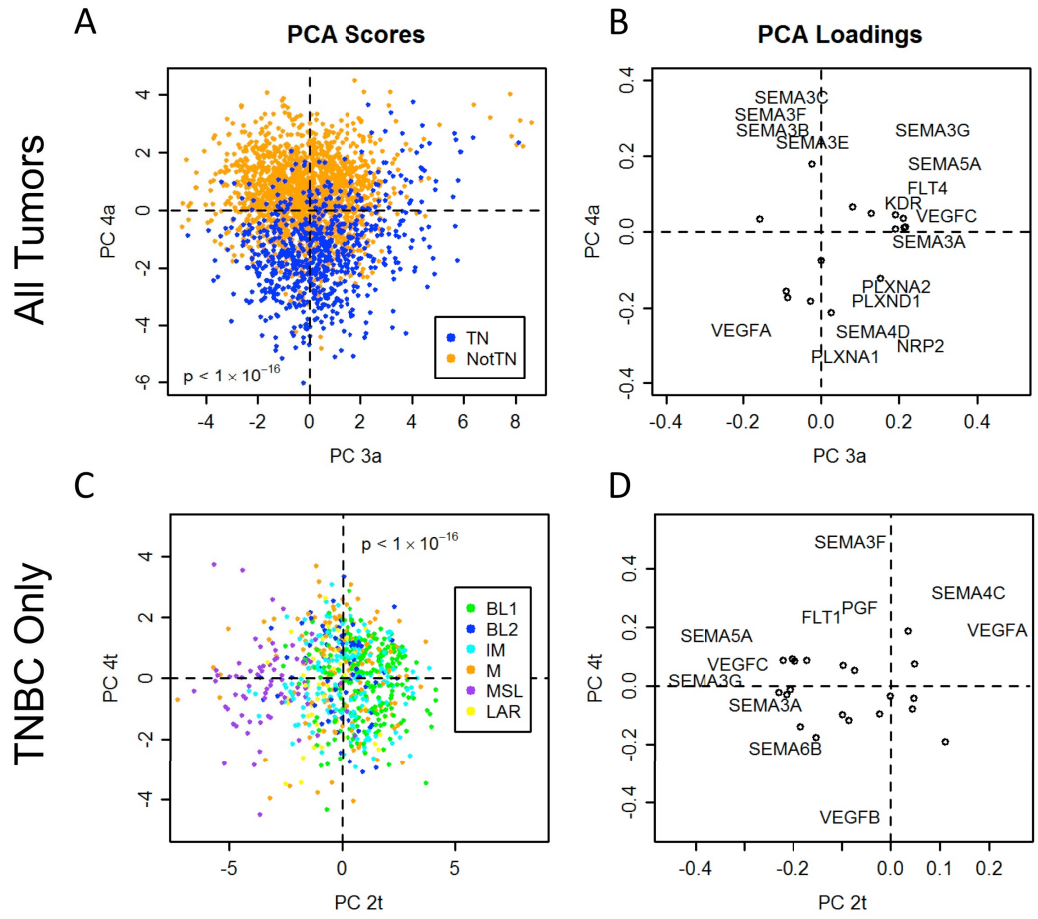

Figure 2-4: Triple negative breast cancers and the mesenchymal stem-like (MSL) subtype of triple negative breast cancers are associated with increased pro-angiogenic gene expression and decreased anti-angiogenic gene expression.

A-B, PCA scores (A) and loadings (B) for VEGF- and semaphorin-related genes in all tumors. Gene names are only shown for probes whose loadings on the two principal components exceeded a radius of 0.2 from the origin. Circles denote genes whose names do not appear. Triple-negative (TN) samples project to lower values of tumor PC4a, corresponding to high VEGFA and low SEMA3 expression. C-D, PCA scores (C) and loadings (D) for VEGF- and semaphorin-related genes in only the TN samples. The MSL subtype projected to low values of TNBC PC2t, corresponding to up-regulation of VEGFC, SEMA3G, and SEMA5A and down-regulation of VEGFA.

significantly associated with the basal subtype (as defined by the PAM50 gene signature classifier described previously [49]) (data not shown), consistent with the similarities between TNBCs and the basal subtype.

Low PC4a scores were associated with a pro-angiogenic signature consisting of high expression of VEGFA, SEMA4D, NRP2, and PLXNA1 and low expression of SEMA3B, SEMA3C, SEMA3E, SEMA3F, and SEMA3G. High PC3a scores were associated with high expression of VEGFC, SEMA3A, SEMA3G, SEMA5A, KDR, and FLT4, and low expression of VEGFA (**Figure 2-4B**). In addition to the established roles of VEGFA and VEGFC as promoters of angiogenesis, published experimental data has shown that other genes associated with low PC4a and high PC3a, SEMA4D and SEMA5A, also have pro-angiogenic function [14,15,91]. Interestingly, three of the ligands with reduced expression in high PC3a and low PC4a samples, SEMA3B, SEMA3F, and SEMA3G, had both anti-angiogenic and tumor suppressor functions [11,12,83,88,89,90]. The role of SEMA3C in angiogenesis has not been well-defined, but like other class-3 semaphorins, it binds to neuropilin receptors. Thus, it may impair signaling by members of the VEGF family by competing for neuropilin.

Examining the correlations between PC4a scores and all genes whose expression was measured on the U133A platform revealed that the ESR1 gene, which encodes ER, had the second highest correlation with PC4a of all genes (data not shown). Other transcription factors associated with ER, such as GATA3 and FOXA1, had high correlations as well. This indicated that the association between PC4a score and TN status may arise primarily because of an association with ER, as opposed to PR or HER2. ER, PR, and HER2 did not appear in the list of the most correlated genes with PC3a scores (data not shown).

2.3.3 The MSL subtype differs significantly from other TNBC subtypes

Next we examined VEGF and semaphorin expression in TNBC samples assigned to the TNBC subtypes discovered in Lehmann et al [81]. PCA of VEGF and semaphorin expression for

only the TNBC samples revealed that of all of the subtypes, the mesenchymal stem-like (MSL) subtype was most distinguishable from the others (**Figure 2-4C**). The MSL subtype projected to low values of the second principal component (PC2t, "t" to denote the TNBC-only data set). The gene expression pattern corresponding to low PC2t included low expression of VEGFA and high expression of VEGFC, SEMA5A, and SEMA3G (**Figure 2-4D**). This was similar to the PC3a from the analysis of all tumors in the previous section (**Figure 2-4B**), except that the signs were reversed. There was substantial overlap between the triple negative tumors that had high PC3a scores in Figure 2-4A and the tumors that had low PC2t scores in Figure 2-4C, indicating that the MSL subtype could likely be distinguished even in the PCA of all of the tumors.

The median expression of VEGFA, VEGFC, SEMA5A, and SEMA3G in the MSL subtype was closer to the median of receptor-positive tumors than to that of TNBC samples. In the case of VEGFA, SEMA5A, and SEMA3G, the expression levels in the MSL subtype were closer to those of normal tissues than the tumor average. On the other hand, VEGFC was expressed at higher levels on average than in any of the other groups, indicating that this subgroup may be susceptible to angiogenesis inhibitors that target VEGFC instead of VEGFA.

2.3.4 Consensus clustering defines VEGF- and semaphorin-based tumor subtypes

We used consensus *K*-means clustering to determine VEGF- and semaphorin-related subtypes independent of any other classifications. This differed from the PCA-based clusters in **Figure 2-4** in that all gene expression variation was analyzed here to determine the natural clusters that arise in VEGF- and semaphorin-related gene expression. Consensus clustering revealed 7 clusters (labeled A-G) for "all tumors" data (data not shown), and 5 clusters (labeled J-N) for "TNBC-only" data (data not shown). The gene expression for the seven tumor clusters is illustrated in the heatmap in **Figure 2-5**, with the clusters arranged by the number of TNBCs present in decreasing order.

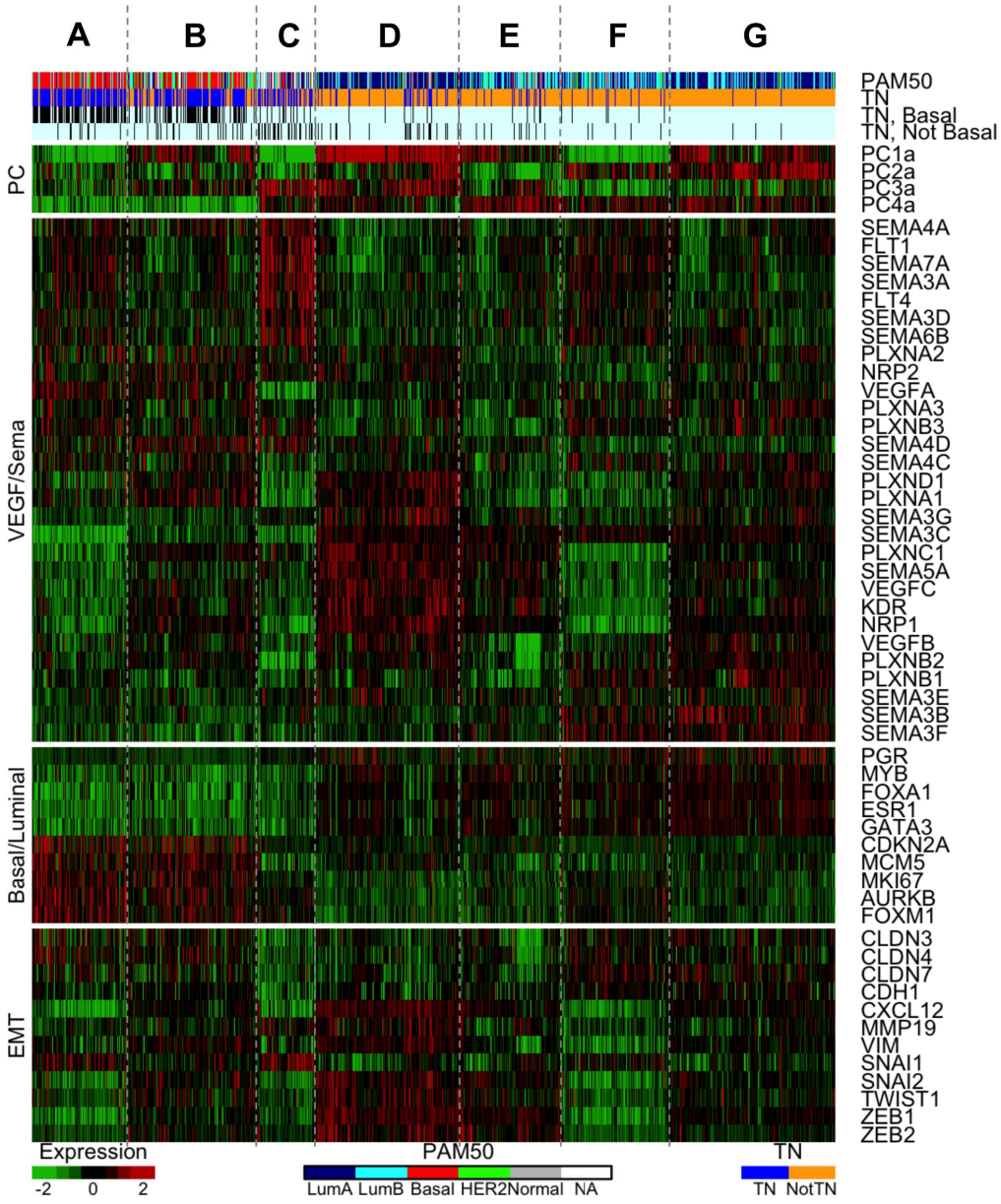

Figure 2-5: Heatmap of the 7 VEGF/Sema-based tumor clusters.

Samples are ordered across the columns by cluster membership as determined by consensus *K*-means clustering. The clusters are ordered by TN content, with cluster A on the left having the highest percentage of TNBCs. The VEGF/Sema-based clusters are able to differentiate the basal intrinsic subtype from the other intrinsic (PAM50) subtypes. As previously noted, the basal subtype is strongly associated with triple negative tumors (third and fourth bars: black for basal

TNBCs and non-basal TNBCs, respectively). Other breast cancer-related genes have expression patterns that align with the VEGF/Sema-based clusters.

High VEGFA-expressing clusters. The first two clusters (A and B) in Figure 2-4 possessed the pro-angiogenic PC4a gene expression signature noted in Figure 2-3, namely high VEGFA expression and low expression of SEMA3B/3C/3F. Cluster A is distinguished from cluster B by higher expression of FLT1 (VEGFR1), FLT4 (VEGFR3), and several semaphorins (including SEMA3A), and lower expression of VEGFC, KDR (VEGFR2), and NRP1. Both clusters have high percentages of TNBCs (78% and 66%, respectively). Using the intrinsic classifier (PAM50, Figure 2-4), these two clusters were found to contain most of the basal subtype tumors. As expected, most of the TN tumors were basal as well (rows 3 and 4 of Figure 2-4 compare the overlap between the TN and basal subtypes). Both of these clusters had low expression of the genes encoding for ER and PR (ESR1 and PGR) and of some of their associated transcription factors (GATA3, FOXA1, MYB), and high expression of proliferation-related genes (the basal/luminal panel in **Figure 2-5**), consistent with the basal subtype.

High VEGFR1/VEGFR3-expressing cluster. Cluster C in **Figure 2-5** had high expression of FLT1 (VEGFR1), FLT4 (VEGFR3), SEMA3A, and some other semaphorins, with low expression of VEGFA, VEGFC, KDR (VEGFR2), NRP1, SEMA3C, PLXNA1, and PLXND1. This cluster also had a relatively high percentage of TNBCs (43%), but with a lower amount of basal subtype tumors than the high VEGFA-expressing clusters. This indicated that although many TNBCs had the pro-angiogenic PC4a signature, it was not strictly required for a tumor to be triple-negative. This cluster had low expression of the claudin genes CLDN3, CLDN4, and CLDN7, raising the possibility that tumors in this cluster were members of the claudin-low subtype, a group of breast tumors known for their invasive, mesenchymal-like behavior.

High VEGFC-expressing cluster. Cluster D in **Figure 2-5** had high expression of a group of genes including VEGFC, KDR (VEGFR2), NRP1, and SEMA5A. This corresponded to the alternative pro-angiogenic TNBC PC2t signature noted in **Figure 2-4**. This cluster had some TNBCs (16%). Significant overlap was noted between this cluster and the luminal A subtype of the PAM50 intrinsic classifier; 67% of cluster D tumors were luminal A, representing 30% of all luminal A tumors. Cluster D was notable in that it had the highest expression of transcription factors implicated in the epithelial-to-mesenchymal transition (EMT), including SNAI2, TWIST1, ZEB1, and ZEB2 (the panel labeled "EMT" in **Figure 2-5**). This could indicate a role for VEGFC-mediated signaling in tumors undergoing an EMT.

High SEMA3-expressing clusters. Clusters F and G in **Figure 2-5** had the anti-angiogenic high PC4a signature described previously: high expression of the anti-angiogenic semaphorins SEMA3B, SEMA3E, and SEMA3F, with low expression of VEGFA. These clusters had the lowest number of TNBCs and were mostly luminal A or B when classified into the PAM50 intrinsic subtypes. The pattern of expression of luminal markers and proliferation-related genes was opposite to that noted for the high VEGFA-expressing clusters: expression of ESR1, PGR, and associated transcription factors was high while expression of proliferation-related genes was low.

2.3.5 Consensus clustering defines VEGF- and semaphorin-based TNBC subtypes

The 5 TNBC clusters, denoted J-N (**Figure 2-6**), were ordered as closely as possible to the tumor clusters in **Figure 2-5**. Thus the majority of samples in cluster A of **Figure 2-5** fall into cluster J of **Figure 2-6**, and so on. The relationship is not perfect; many samples are differentially classified between the two cluster analyses.

The five TNBC clusters had similar expression patterns to the clusters described above for all tumors. TNBC clusters J and K had higher VEGFA expression on average, and as with clusters A and B, were differentiated by the pattern of high FLT1, FLT4, SEMA3A in cluster J

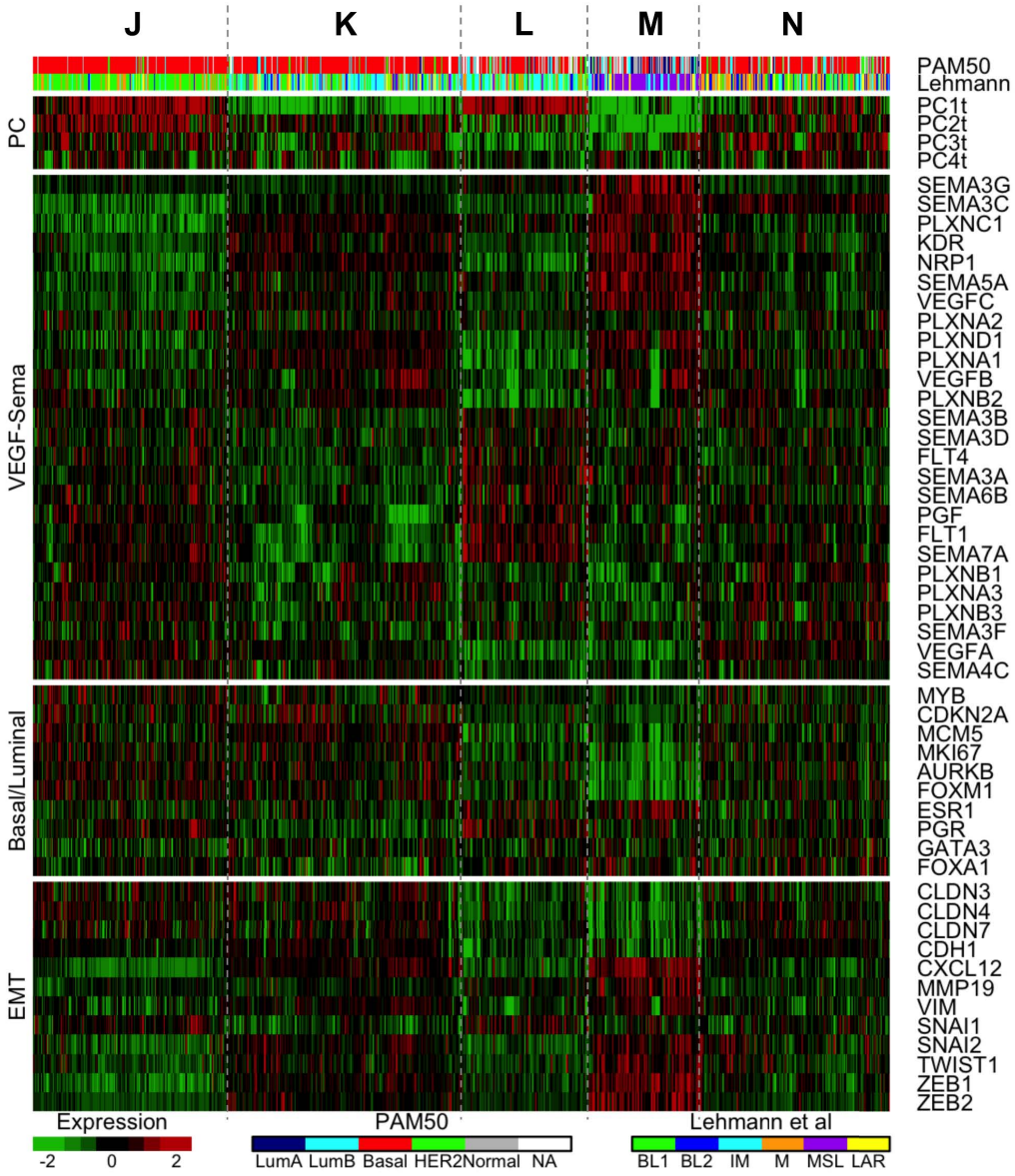

Figure 2-6: Heatmap of the 5 VEGF/Sema-based TNBC clusters.

Consensus *K*-means clusters of only the TNBC data are arranged according to the ordering from Figure 2-5, in that the TNBCs that appeared in tumor cluster A generally now appear in TNBC cluster J (although the correspondence is not perfect) and so on for the other clusters. The VEGF/Sema-based clusters are able to differentiate the MSL subtype from Lehmann et al. (cluster M) from the other TNBC subtypes. Claudin-low subtype-related patterns of gene expression in the panel labeled "EMT" were associated with cluster M, the MSL-enriched cluster.

and high KDR, NRP1 in cluster K. Clusters L and M had lower VEGFA expression, with cluster L expressing high levels of PGF, FLT1, FLT4, and SEMA3A, and cluster M expressing high levels of VEGFC, KDR, and NRP1. Cluster D had low PC2t scores from **Figure 2-4C/D**. The fifth cluster was not particularly distinguishable from the other TNBC clusters. Notably, all the clusters except cluster M were heavily populated by tumors of the basal PAM50 intrinsic subtype. This was seen for clusters A and B in the all-tumor data set (**Figure 2-5**), but not for clusters C and E. This is further evidence for the association between the basal subtype and TNBC; basal tumors comprised a small minority of clusters C and E in **Figure 2-5**, but this minority became the majority in **Figure 2-6** when only the TN tumors were considered.

The TNBC subtypes found in Lehmann et al [81] had some associations with the VEGF-/Sema-based clusters found here. Of tumors in the MSL subtype, 81% were found in cluster M, comprising 65% of the tumors in that cluster. This corresponds to the PCA results that demonstrated a strong association of PC2t with the MSL subtype. Most of the other subtypes were evenly distributed across the clusters, with the exception of the basal-like 1 (BL1) subtype, which comprised 66% of cluster J.

Patterns of expression for other genes when sorted in the order of the TNBC clusters were less apparent in **Figure 2-6** than in **Figure 2-5**, with cluster M showing the most significant regulation. Growth-associated genes such as FOXM1, AURKB, and MKI67 were strongly down-regulated in this cluster while the EMT-associated transcription factors SNAI2, TWIST1, ZEB1, and ZEB2 were up-regulated. Notably, the claudin genes CLDN3, CLDN4, and CLDN7 were down-regulated the most in this cluster, consistent with the observation that the MSL subtype and claudin-low subtype are closely related [97].

2.3.6 Validation using TCGA data

Using two TCGA data sets consisting of 537 tumors quantified using a different microarray platform (**Figures 2-7A** and **2-7B**) and 750 tumors quantified using an RNA-Seq

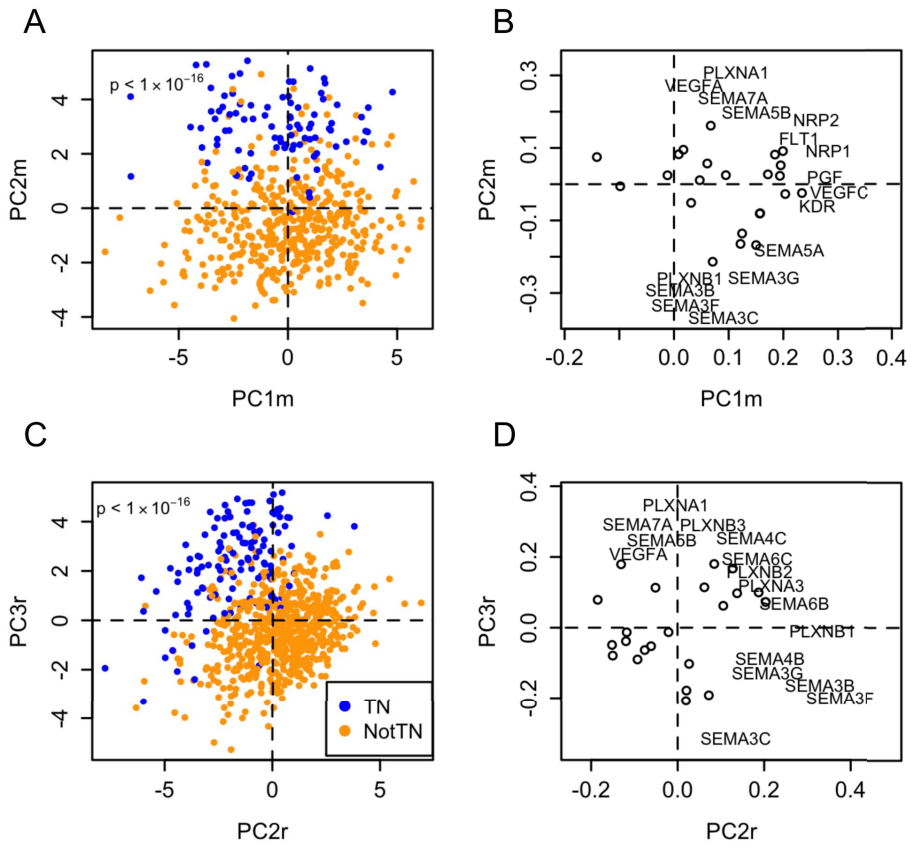

Figure 2-7: Validation of VEGF- and Semaphorin-related gene expression differences between triple-negative and non-TN breast cancers using TCGA data.

A-B, Principal component analysis of TCGA microarray data set consisting of 537 breast tumors. **C-D**, Principal component analysis of TCGA RNA-Seq data set consisting of 750 tumors. Both data sets were processed to gene-level measurements (rather than probe-level) prior to downloading. Tumors were classified as triple negative based on gene expression data for the ESR1, PGR, and ERBB2 genes as described in the *Methods*, all of which had clear bimodal distributions.

platform (**Figures 2-7C** and **2-7D**), we showed that the same patterns of gene expression that distinguish TNBCs from other tumors could be found in other patients using different technologies. PCA scores between the two platforms used in the TCGA datasets had strong correlations (data not shown). Patterns of gene expression associated with the 2,656-tumor dataset were found in both TCGA datasets (data not shown), including the low PC4a gene expression signature of high VEGFA and low SEMA3B, SEMA3C, SEMA3F, and PLXNB1. Some additional genes were altered consistently in the two validation data sets as well: SEMA5B, SEMA7A, and PLXNA1. All three of these had similar expression patterns to that of VEGFA.

2.3.7 Survival analysis of clusters

We performed Kaplan Meier survival analysis on the tumors to determine the impact of the PCA-derived clusters on patient prognosis. Triple negative status and increasing stage of the tumor were both correlated with poorer prognoses as expected (**Figures 2-8A** and **2-8B**). Multivariate survival analysis using a Cox proportional hazards model showed that tumor stage, lymph node status, PC3a score, and PC4a score were all independent prognostic factors. Interestingly, triple negative status, which was clearly correlated with poor survival, was not significant in the multivariate model ($p=0.06$). The likely reason for this is that TN status is also highly correlated with PC4a scores. There were more non-TNBC samples with low PC4a scores ($n=89$) than TNBC samples with high PC4a scores ($n=29$), possibly resulting in a stronger survival effect from PC4a score than from TN status (**Figure 2-8C**). We also examined the interaction of ESR1 expression with PC4a; both low ESR1 expression and low PC4a scores were significantly associated with poor prognoses (**Figure 2-9**). Interestingly, PC4a score was significantly associated with survival in a subgroup consisting of tumors with high ESR1 expression. Although ER+ tumors are already associated with effective therapies, combination of existing therapies with angiogenesis inhibition may provide additional benefits for these low PC4a, ER+ tumors.

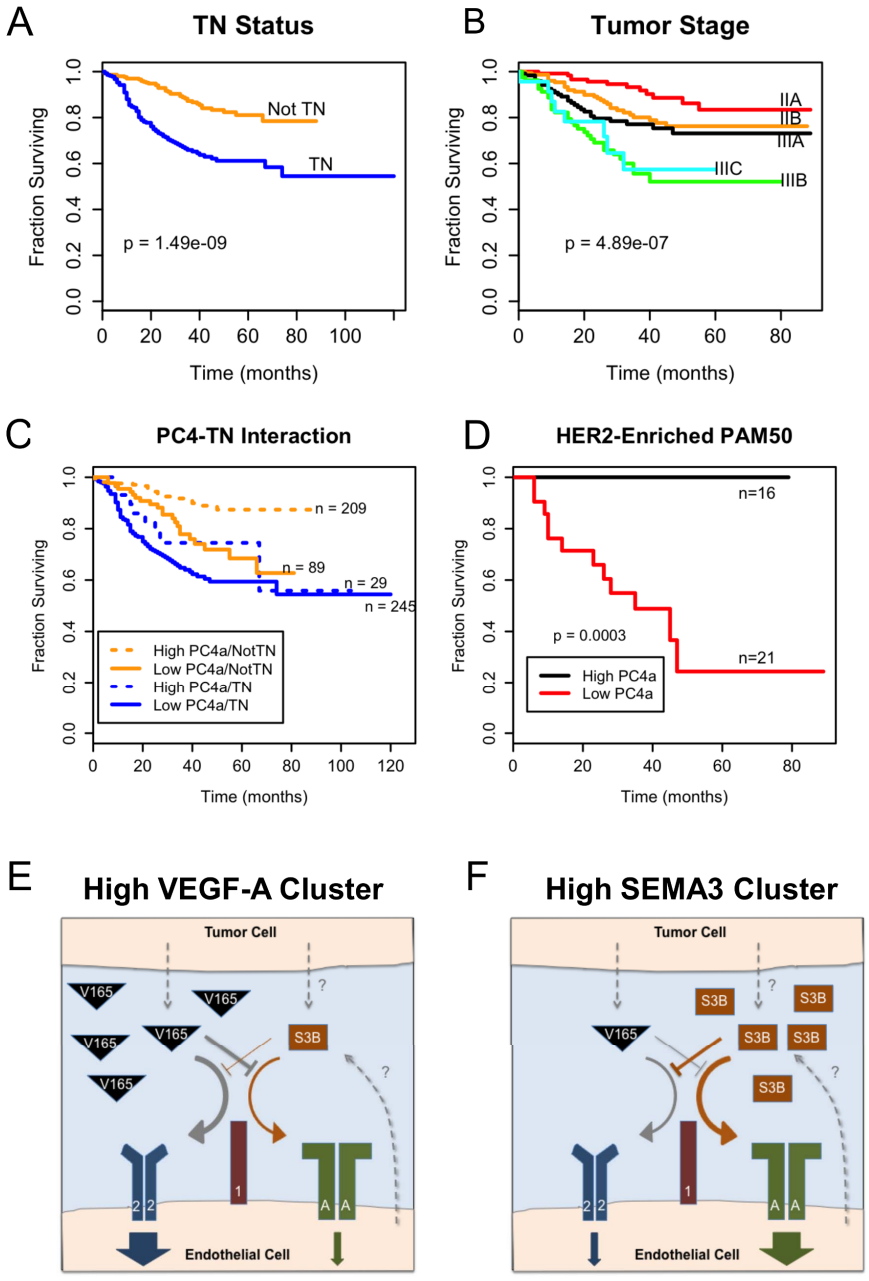

Figure 2-8: Angiogenesis gene expression subgroups correlate with survival.

A-D, Survival curves for tumor samples that had available survival data. A log-rank test was used to determine p-values. Triple negative (TN) status was associated with worse prognosis for 572 patients with survival data (A). Higher tumor stage correlated with worse prognosis in 508

patients with available stage and survival data (B). Non-TNBC patients with low PC4a scores had poor prognoses similar to the TNBC patients (C). Patients in the HER2-enriched PAM50 subtype had significantly poorer prognoses if they also had low PC4a scores (D). **E-F,** Schematic of VEGF/semaphorin competition in the tumor microenvironment. The gene expression patterns of the different subgroups of TNBC and other cancers suggest different regulation of pro- and anti-angiogenesis pathways. The case with high expression of VEGFA and low SEMA3B, which corresponds to the high PC4a group, results in increased signaling through VEGF receptors such as VEGFR2 (blue). Most TNBC fit this profile, although many non-TNBC did also and these showed decreased 5-year survival similar to TNBC (C). Lower expression of VEGFA with high SEMA3B, corresponding to the low PC4a group, results in reduced signaling through VEGF receptors and more signaling through semaphorin receptors such as PLXNA1 (green). Note that these schematics only consider receptor expression.

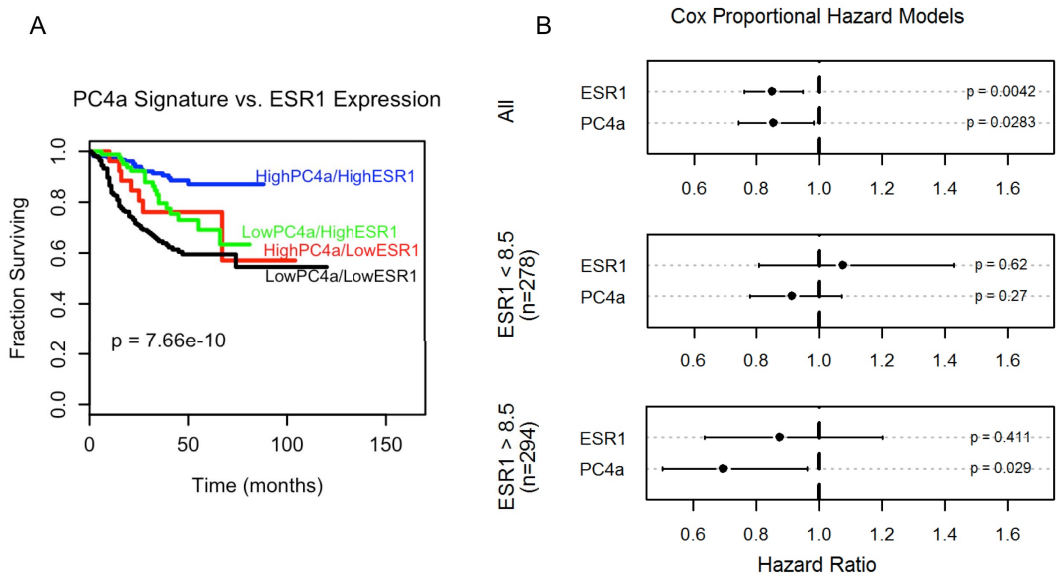

Figure 2-9: Survival analysis based on PC4a and ESR1 expression.

A, Low PC4a scores and low ESR1 expression ("LowPC4a/LowESR1") were both associated with poorer prognoses. In high ESR1-expressing tumors ("HighESR1"), low PC4a scores resulted in poor prognosis as well**,** while high PC4a scores resulted in significantly better prognoses. **B,** Cox proportional hazard models reinforced the overall independence of ESR1 expression and PC4a score. Both factors were significantly associated with survival in a model of all tumors with survival data (top panel). In tumors with low ESR1 expression (middle panel), neither factor was significant. In a model of tumors with high ESR1 expression (bottom panel), PC4a score, but not ESR1 expression, had significant association with survival.

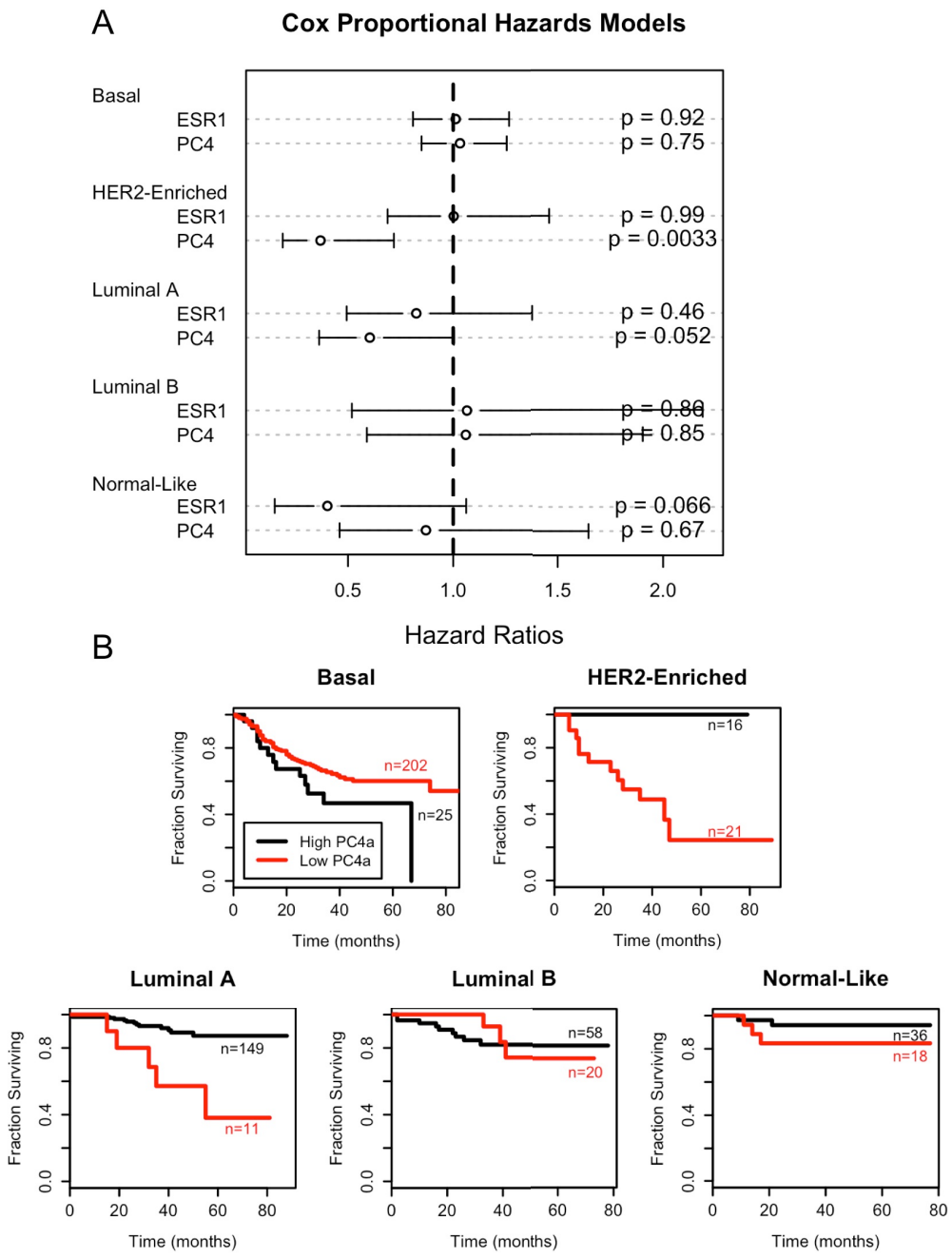

Figure 2-10: Association between PC4a score in the five PAM50 subtypes.

A, Cox proportional hazard models for each PAM50 subtype demonstrated that PC4a score only was significantly associated with survival in the HER2-enriched subtype, while ESR1 expression was not significantly associated with survival in any of the subtypes. Hazard ratios indicate the effect of increasing PC4a score or ESR1 expression; thus, a lower hazard ratio indicates that high

PC4a scores are associated with improved prognosis. PC4a scores and ESR1 expression were included as continuous variables. **B,** PC4a scores below the median were typically associated with poorer prognoses except in the basal subtype. As the basal subtype is associated with low PC4a scores, very few samples in the basal subtype had high PC4a scores. This low number of samples explains the steep drop-off in the upper left plot.

Survival analysis by cluster (data not shown) showed that tumor clusters F and G had significantly better outcomes than the rest of the clusters. These were the only clusters that had both low PC3a scores and high PC4a scores (both anti-angiogenic signatures), reinforcing the prognostic value of these two principal components. The five TNBC clusters found here did not have significantly different prognoses; instead the survival curves of the clusters shared the same poor prognosis characteristic of TNBCs (data not shown). Despite the lack of variability, the differences in patterns of VEGF and semaphorin gene expression may indicate different growth factor dependencies. For example, VEGFC-targeting therapies may be more effective in cluster M, while the rest may benefit more from VEGFA-targeting therapies.

Within each PAM50 subtype, ESR1 was not significant (**Figure 2-10A**) while PC4a score was significantly associated with survival only in the HER2-enriched subtype (**Figures 2-8D** and **2-10B**). The lack of association between survival and PC4a in the basal and luminal PAM50 groups is not unexpected since PC4a is somewhat consistent within these groups (low in basal, high in luminal, as shown in the heatmap in **Figure 2-5**). The prognostic effects of PC4a score in the HER2-enriched subtype could indicate that a low-PC4a subgroup of patients treated with HER2-targeting therapy may benefit from the addition of an anti-angiogenic drug to their treatment. Overall the association of survival and PC4a score in particular subtypes may aid in selecting patients where anti-angiogenic therapy would provide the greatest benefit.

2.4 Discussion

Just as individuals have distinct genomic and gene expression profiles, so too the tumors of each individual are distinct. Understanding and quantifying this variability and individuality is crucial for the development and targeting of therapeutics for diseases as complex and heterogeneous as cancer. Triple negative breast cancers (TNBCs), in particular, are a diverse and difficult-to-treat set of tumors defined primarily by molecular targets for treatment that they do not express, rather than targets that they do express. Angiogenesis, a blood vessel morphogenesis

process underpinning the growth and metastasis of most tumors, is a possible common target for TNBCs, and vascular endothelial growth factor (VEGF) has been targeted in breast cancer as a key regulator of angiogenesis. However, this has succeeded only for a subset of breast cancer patients, and thus understanding which subsets of patients may be responsive to this treatment is desirable. This requires data from a large number of patients, and we used one type of patient population data, gene expression microarrays, to quantify changes in VEGF and semaphorin expression to define relevant patient subgroups.

A high proportion of the genes considered here were significantly different between normal breast tissue and breast tumors (34/55 probe sets with $p<0.001$). This high rate of significance may be an indicator that these genes are heavily regulated by the genomic changes that occur in tumors. Alternatively, the low number of normal samples (n=42) relative to tumor samples (n=2,656) may result in an unrealistic estimate for significance. It is important to note that the range/variability in tumor expression is very high compared to the normal samples, as indicated by the standard deviations of each group (data not shown). Although the mean values of expression for tumors and normal tissues may differ, the range of tumor expression often overlaps the range of normal expression. This makes individual genes poor biomarkers; however, they can be combined to identify tumor subgroups that correlate with differences in tumor characteristics.

The overall expression changes were consistent with previously reported breast cancer data of these genes at the mRNA and protein level. The increased expression of VEGFA in TNBCs compared to non-TNBCs was consistent with previous work that found an approximately 3-fold increase of VEGF as measured by ELISA of intra-tumoral samples from 679 patients [98]. VEGFR2 (KDR) was previously found to be significantly associated with TNBC in a panel of tissue microarrays from 564 patients [98]. This is consistent with our results, which showed a relatively high loading of KDR on the principal component associated with VEGFC, NRP1, and PLXND1, which had the second highest association with triple-negative (TN) status in the all-

tumor data set. The increased expression of VEGFC in TNBC samples found here has also been demonstrated in IHC of breast cancer sections, where VEGFC stained positively in TNBCs significantly more often than in non-TNBCs [99]. Studies examining the amount of semaphorins and plexins expressed in breast cancer patients by TNBC and non-TNBC subgroups are not available, but there are some reports comparing their expression in normal breast and tumor tissue. SEMA3A, SEMA3B, SEMA3F, PLXNA1, and PLXNA3 were all shown by IHC to decrease as tumors progressed, while NRP1 increased and NRP2 stayed the same [100]. Another study showed the same pattern for PLXNA3 expression, while also showing that SEMA4F expression increased as tumors progressed [101].

Using multiple clustering and analysis algorithms, we have revealed patterns of gene expression associated with the triple-negative subtype of breast cancer that may indicate a higher collective pro-angiogenesis activity (**Figure 2-5**, tumor clusters A-C). In addition to up-regulation of the well-known angiogenic growth factor VEGFA in clusters A and B, several anti-angiogenic semaphorins were down-regulated. SEMA3B and SEMA3F have both anti-tumorigenic and anti-angiogenic properties [11,12,83,88,89,102,103]. SEMA3C, whose role in angiogenesis is less well-understood [84,85], was also consistently down-regulated in tumors with high expression of VEGFA. Given that class-3 semaphorins compete with VEGF for binding to neuropilin co-receptors, this overall pattern of gene expression may enhance VEGF signaling in three ways: directly by increasing the amount of VEGFA; indirectly by reducing the amount of competitive inhibition for neuropilin; and decreasing anti-angiogenic Sema-Plexin signaling. The survival analysis in **Figure 2-8C** demonstrates the significance of this high-VEGF/low-Sema3 signature; patients with this signature have similar poor prognoses regardless of their triple negative status. The activity of angiogenesis in the TNBC-enriched cluster C from **Figure 2-5** is less clear: VEGFA is down-regulated, while several semaphorins with both pro-angiogenic (SEMA6B and SEMA7A) and anti-angiogenic (SEMA3A and SEMA4A) effects are up-regulated.

What do these different gene-expression subgroups mean for treatment? We would hypothesize that high-VEGFA-expressing tumors would be more vulnerable to anti-VEGF treatment. Clinical trial results for the anti-VEGF drug bevacizumab have thus far not shown an increased efficacy in triple-negative subtypes [29,76,104,105,106]; instead, similar improvements have been seen in both triple negative and hormone receptor positive cancers (all cases were HER2-negative). However, we note that the high-VEGFA, low-Sema3 pattern (clusters A and B in **Figure 2-5**) makes up only 69% of all TNBC samples, and that 12% of non-TNBC samples can be classified as having a similar gene expression profile, possibly confounding this analysis.

Other clusters found in the tumor data may be less susceptible to inhibition of angiogenesis. The clusters with high expression of class-3 semaphorins would likely not benefit from this type of therapy because class-3 semaphorins function as endogenous inhibitors of angiogenesis. These tumors would be expected to be less aggressive; survival curves for patients with high class-3 semaphorin expression have the best prognoses (high PC4a in **Figure 2-8C**). The gene expression pattern consisting of high expression of VEGFC, PlGF, NRP1, and PLXND1 and low class-3 semaphorin expression (PC3a in **Figure 4-4A/B**) is also likely to be pro-angiogenic. However, it would not be expected to benefit from an anti-VEGFA therapy; instead, a different target would be needed to inhibit angiogenesis. This is supported by a report that low IHC staining of VEGFC and NRP1 is associated with improved progression-free survival in patients receiving bevacizumab, while the level of VEGFA was not associated with changes in progression-free survival [107].

The TNBC subtypes previously identified [81] demonstrated similar expression of VEGF- and semaphorin-related genes with the exception of the mesenchymal stem-like subtype. This subtype was noted for its enrichment of genes involved with migration and growth factor pathways, including KDR [81]. Here, we found a cluster of angiogenesis-related genes with increased expression in the MSL subtype, including VEGFC and KDR (TNBC cluster M in

Figure 2-6, corresponding to tumor cluster D in **Figure 5-5**). Notably, however, VEGFA expression was decreased, indicating that although angiogenesis may occur in tumors of this subtype, VEGFA-targeted therapies are not likely to be successful inhibitors. In the analysis of all tumors, this VEGFC-dominated signature (tumor cluster D) was present in 18.5% of tumors. This cluster had a low proportion of triple negative tumors, raising the possibility that the MSL subtype may not just be a small subgroup within TNBCs, but a therapeutically relevant subgroup of breast cancers as a whole.

The concordance of the VEGF-/Sema-based clusters that we found here with expression patterns of genes associated with the basal/luminal distinction and EMT suggests that different breast cancer subtypes utilize the VEGF and semaphorin signaling pathways in consistently different ways. In particular, basal tumors with high expression of growth-associated genes such as MKI67 and AURKB tend to have higher levels of VEGFA, presumably to provide the rapidly proliferating cells with sufficient vasculature. On the other hand, tumors with low expression of growth-associated genes but high expression of EMT-associated transcription factors such as SNAI2 and TWIST1 have low VEGFA expression and high VEGFC expression. The lymphangiogenic VEGFC may facilitate invasion by allowing tumor cells to travel through the lymphatics, a commonly used route of metastasis in breast cancer [99]. This highlights the usefulness of this study not just in targeting anti-angiogenic therapies, but in understanding tumor biology as well.

One limitation of using gene expression microarrays on tumor samples taken from biopsies or surgeries is that the samples are heterogeneous. Along with the tumor cells they also contain stromal cells, including endothelial cells, fibroblasts, and immune cells. The expression of most of the ligands considered here can be assumed to be predominantly attributable to expression in the tumor cells, but for receptor expression the analysis is less straightforward. This is particularly true for receptors whose primary function of interest is on a cell type making up a

small percentage of the total, e.g. endothelial cells. Their expression may be up-regulated in those cells but down-regulated in the more numerous cell type, resulting in detection of no or opposite change in expression in the microarray measurement of the heterogeneous sample. Immunohistochemistry can address this issue by measuring the cell-type-specific protein expression. For example, studies in a wide range of breast tumors have shown that NRP1 and NRP2 are both expressed on almost all endothelial cells, but very rarely on breast tumor cells [108,109]. Conversely, PLXNB1 has been shown to be expressed on the surface of tumor cells, but less so on neighboring endothelial cells [110]. Thus, differences in expression of NRP1 and NRP2 measured by microarray can be assumed to be primarily due to endothelial cells, and differences in PLXNB1 due to tumor cells. Laser capture microdissection or other sorting methods could also resolve cell type differences by isolating specific cell types prior to analyzing gene expression. Methods such as these will be particularly useful in determining the relative amount of VEGF signaling taking place in tumor and endothelial cells.

We have yet to determine whether the VEGF- and Semaphorin-based clusters found here are recapitulated in gene expression data for breast cancer cell lines. Extensive work has been done to characterize the subtypes found in these cell lines and the differential susceptibility of the cell lines to various therapeutics [111,112]. Many aspects of VEGF and semaphorin signaling depend on other cells in the tumor microenvironment, in particular endothelial cells and tumor stromal cells, and an analysis of cell lines could aid in determining which differentially expressed ligands in the present study arise due to tumor cells and which are due to stromal cells; as well as insight into whether observed receptor expression variation is due to tumor cells or tumor-associated endothelial cells.

We analyzed survival data in part to assess whether the VEGF- and Semaphorin-based clusters were associated with prognosis of breast cancer patients. The high correlation of VEGF- and Semaphorin-related gene expression with existing prognostic indicators such as TN status

confounds the analysis and makes it impossible to determine from this data why some patients have poorer prognoses. However, we used a multivariate Cox proportional hazards model and Kaplan-Meier plots to demonstrate that a subgroup of ER+ tumors with the pro-angiogenic PC4a signature had poorer prognosis. Thus, the pro-angiogenic PC4a signature may have a role in severity of the disease, independent of ER or TN status. To determine the actual significance of the VEGF- and Semaphorin-based groups found here, experimental models of breast tumors are needed. Tumor xenografts in immunocompromised mice could be used to measure the growth and invasion of tumors of the various subtypes. This type of experimental model provides the advantage of allowing for other processes that contribute to cancer progression other than tumor cell growth, including angiogenesis. Targeted VEGF inhibitors and inhibitors of VEGF-pathway receptors could be administered to show whether the VEGF-/Semaphorin-based signature found here is truly relevant in tumorigenesis.

Computational models of VEGF and Semaphorin ligand-receptor interactions will be useful in unraveling the effects of the expression changes found here. The large number of proteins involved, combined with the complexity of their interactions, will make it necessary to use models to understand the overall effect of the expression patterns on signaling through VEGF receptors. Models of VEGF signaling [47,113,114] can be extended to include the Semaphorins found to be relevant in the current study. These models will enable prediction of patients expected to respond to existing therapies and can suggest effective therapeutic targets.

3 VEGF/Semaphorin Expression Subgroups Across Cancer Types

3.1 Summary

Therapies targeting angiogenesis are approved for use in treating multiple types of cancer, but the response rates to these drugs are typically relatively low. Biomarkers that are predictive of response are needed to so that responders can be identified while pursuing alternate therapies for resistant patients. We determined VEGF and Semaphorin patterns of expression in seven cancer types – breast adenocarcinoma, colon adenocarcinoma, glioblastoma, renal cell carcinoma, lung adenocarcinoma, serous ovarian cystadenocarcinoma, and prostate adenocarcinoma – using multivariate statistical methods. Clustering revealed VEGF/Sema subgroups in most types of cancer, some of which overlapped significantly with existing transcriptomic subtypes. Using principal component analysis, we found multivariate patterns of altered VEGF/Sema expression suggestive of pro-angiogenic activity in some subgroups. These VEGF/Sema patterns had prognostic significance in several cancer types, suggesting that these subgroups may be enriched with aggressive tumors in which anti-angiogenic therapies would be of clinical benefit.

3.2 Methods

3.2.1 Data sets

All gene expression data was obtained from the Cancer Genome Atlas (TCGA). Only data from primary tumors was used. The types of cancers included here were: breast adenocarcinoma (*N*=914); colon adenocarcinoma (*N*=192); glioblastoma (*N*=155); renal cell carcinoma (*N*=480); lung adenocarcinoma (*N*=488); serous ovarian cystadenocarcinoma (*N*=262);

and prostate adenocarcinoma (*N*=176). Here, we considered only the expression of the 39 genes encoding ligands and receptors in the VEGF and Semaphorin families.

3.2.2 Transcriptomic subtypes

Subtypes previously defined in TCGA studies are referred to here as "transcriptomic subtypes" because they were defined based on a global clustering analysis of gene expression. Most samples in the TCGA datasets had transcriptomic subtype information available. When subtype information was not available, we built a nearest shrunken centroid classifier using the *pamr* function in *R* using the expression of all of the genes. We chose the shrinkage threshold with 10-fold cross-validation.

3.2.3 Statistical analyses

All methods used in this chapter were described previously: PCA (Chapter 2.2.5), consensus *K*-means cluster (Section 2.2.8), and Kaplan-Meier survival analysis (Chapter 2.2.7).

3.3 Results

3.3.1 VEGF/Sema-based clusters overlap with transcriptomic subtypes

We used consensus *K*-means clustering to group patients according to VEGF/Sema gene expression. We were able to find an optimal number of clusters in almost all cancer types; prostate adenocarcinoma was the one type where consistent partitioning of samples did not occur. We explore prostate cancer further in Chapter 4. For the other cancer types, we found five breast cancer clusters, five colon cancer clusters, four glioblastoma clusters, four renal cell carcinoma clusters, six lung cancer clusters, and five ovarian cancer clusters (labeled at the bottom of each heatmap in **Figure 3-1**).

In most cancer types, at least one VEGF/Sema cluster had significant overlap with a previously defined transcriptomic subtype. As noted in Chapter 2, the VEGF/Sema cluster with high *VEGFA* expression and low class 3 Semaphorin expression (labeled as Cluster 1 in **Figure**

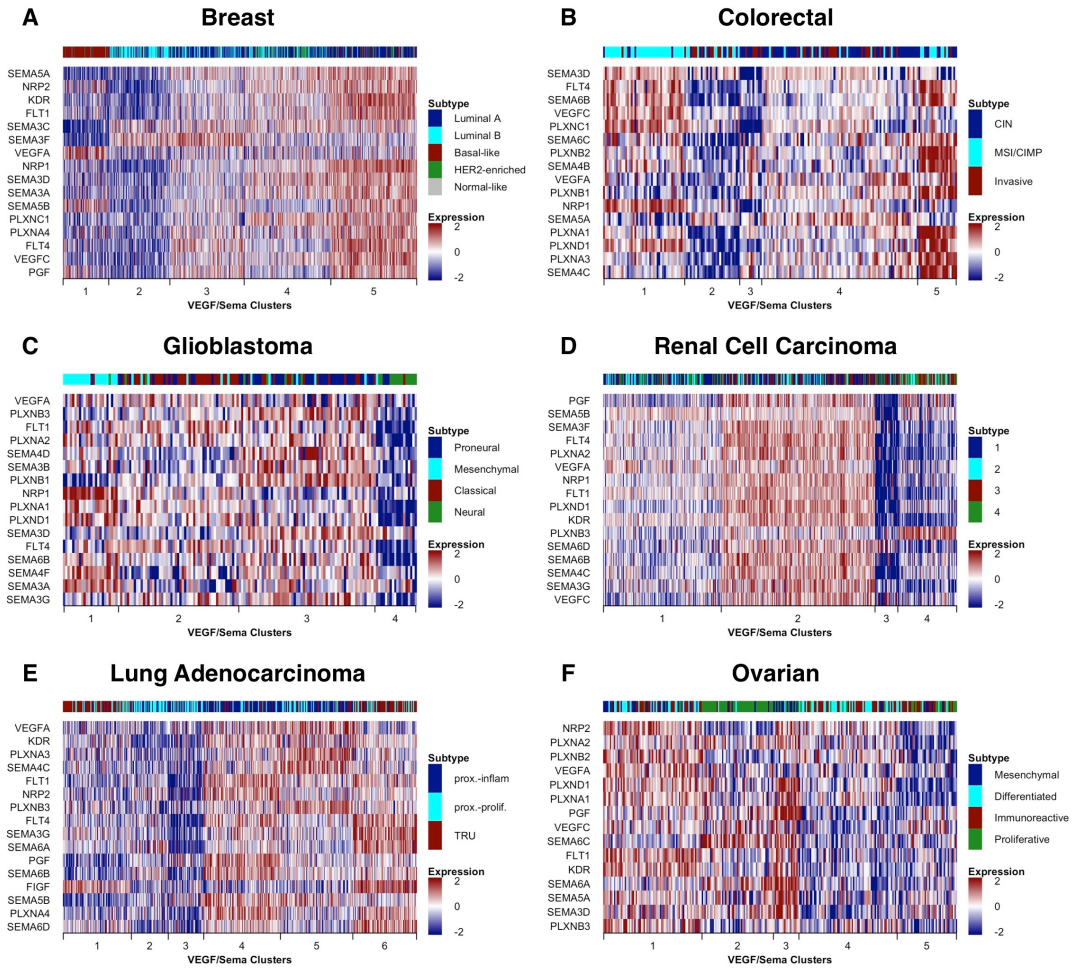

Figure 3-1: VEGF/Sema clustering across cancers.

K-means clustering revealed between four and six clusters in breast adenocarcinoma (A), colon adenocarcinoma (B), glioblastoma (C), renal cell carcinoma (D), lung adenocarcinoma (E), and ovarian cystadenocarcinoma (F). VEGF/Sema expression-based cluster numbers are shown at the bottom of each heatmap, with tick marks separating clusters. Of the 39 VEGF/Sema genes, only the 16 with the highest inter-cluster variance are shown. Colors corresponding to transcriptomic subtypes are shown along the top bar of each heatmap.

3-1A) was strongly associated with the basal-like subgroup, which also significantly overlaps with triple-negative tumors. We assessed the significance of cluster-subtype overlap using Fisher's exact test. The top ten most significant pairs (**Table 3-1**) included pairs from all cancer types except renal cell carcinoma and prostate cancer.

3.3.2 High VEGFA/low Sema3 pattern is unique to triple-negative breast cancer

The pattern of high *VEGFA* expression and low *SEMA3B, SEMA3C, SEMA3D,* and *SEMA3F* expression was observed in basal samples in this data set (**Figure 3-2A**). This was expected as this TCGA data was used in Section 2 as a validation data set. Surprisingly, this pattern of pro-angiogenic gene expression was not observed in any subtype in any other cancer types.

A pro-angiogenic pattern involving up-regulation of *VEGFC*, *KDR*, and *NRP1* was observed in several subtypes. This pattern was present in the MSI-CIMP subtype of colon cancer (**Figure 3-2B**) and partially present in the mesenchymal subtype of glioblastoma (**Figure 3-2C**; *KDR* was not up-regulated in this case). These three genes were also simultaneously down-regulated in the luminal B subtype of breast cancer (PCA plot not shown, but pattern visible in VEGF/Sema cluster 2 in **Figure 3-1A**).

Other patterns of VEGF/Sema expression associated with transcriptomic subtypes lacked pro-angiogenic signatures in the sense that they did not include traditional drivers of angiogenesis such as VEGFA and VEGFC. Nonetheless, some signatures indicated a potential role for angiogenesis: in the ovarian cancer proliferative subtype, PGF was up-regulated, along with SEMA6A and SEMA4F, two semaphorins with links to tumor progression [115,116].

3.3.3 A VEGF/Sema glioblastoma subtype improves prognostic classification

We performed Kaplan-Meier survival analysis using both the VEGF/Sema clusters and the previously-defined transcriptomic subtypes to determine whether the VEGF/Sema clusters improve our ability to separate patients on the basis of overall survival. In most cases, the

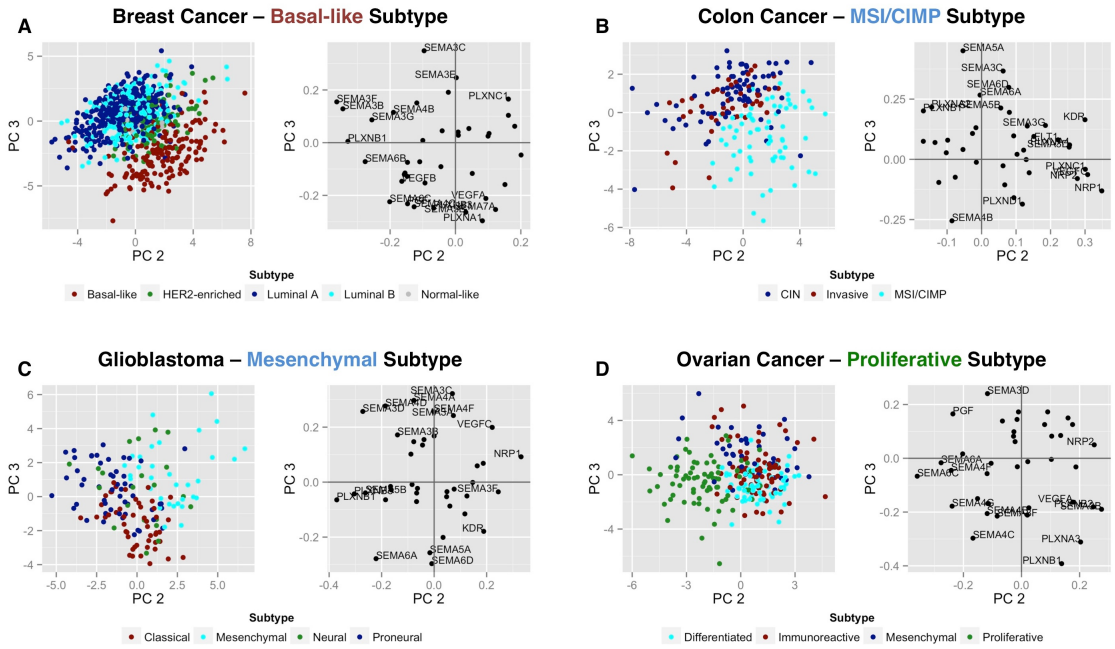

Figure 3-2: Multivariate patterns of VEGF/Sema expression in cancer subtypes.

A-D, PCA of VEGF/Sema expression in breast (A), colon (B), brain (C), and ovarian (D) cancers revealed that some transcriptomic subtypes were easily distinguishable based only on expression of VEGF- and Sema-related genes. For each cancer type, the left panel shows the scores (projection of the samples onto the principal components), while the right panel shows the loadings (projection of genes onto the principal components). Dot colors in the scores plot correspond to transcriptomic subtypes as indicated. Gene names were displayed on the loadings plots if the loading magnitude exceeded a certain threshold.

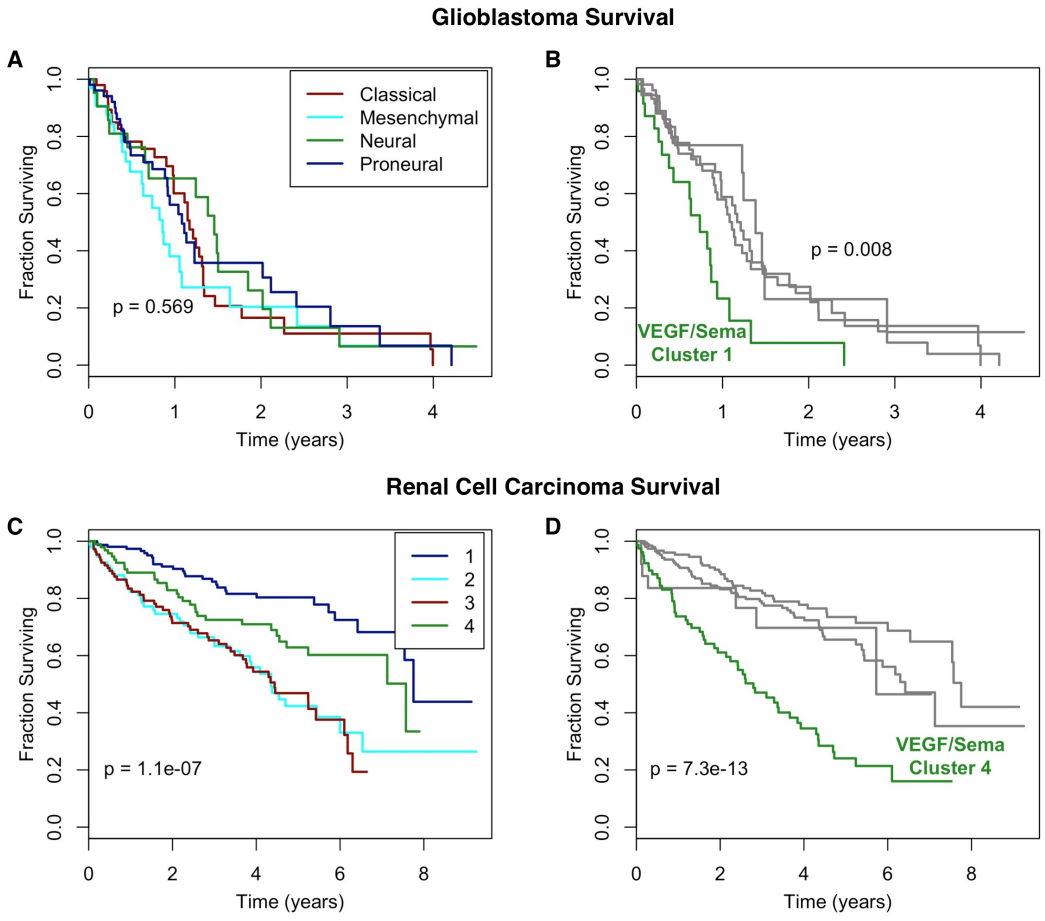

Figure 3-3: VEGF/Sema clusters with prognostic significance.

Whereas transcriptomic subtypes of glioblastoma were unable to differentiate patients on the basis of prognosis (A), VEGF/Sema cluster 1 of glioblastoma (B) did have significantly shorter time to death. The transcriptomic clusters of renal cell carcinoma had prognostic significance (C), but VEGF/Sema cluster 4 (D) was more effective in distinguishing a subset of patients with poor prognosis.

performances of the two groupings were similar: if the transcriptomic subtypes were capable of finding significant survival differences, the VEGF/Sema clusters were as well; on the other hand, if the transcriptomic subtypes did not perform well, the VEGF/Sema cluster also did not.

One exception was VEGF/Sema cluster #1 of the glioblastoma data set, characterized by up-regulation of VEGFC and NRP1. This cluster had significantly worse outcomes than any other cluster (**Figure 3-3B**, P=0.008). The transcriptomic subtypes were unable to separate patients into groups with different outcomes (**Figure 3-3A**, P=0.57). It should be noted that although VEGF/Sema cluster #1 was strongly associated with the mesenchymal subtype, it did not consist exclusively of mesenchymal tumors, nor were all of the mesenchymal subtype tumors in this cluster.

Another cluster with prognostic significance was VEGF/Sema cluster #4 in the renal cell carcinoma dataset (**Figure 3-3D**). Although the transcriptomic subtypes did differ in prognoses (**Figure 3-3C**), the survival curve for VEGF/Sema cluster #4 was lower than any of the curves defined by the transcriptomic subtypes. This VEGF/Sema cluster was associated with high expression of PLXNB3 (**Figure 3-1D**), which can promote angiogenesis through interactions with its ligand SEMA5A [15].

3.4 Discussion

The ability of *K*-means clustering to identify distinct groups based on VEGF/Sema expression in all cancer types (with the exception of prostate) suggests that alterations to VEGF signaling are common across cancer types. Furthermore, based on our analysis, these alterations are expected to be found only in subgroups within each cancer type. Thus, additional research is needed to elucidate the role that particular VEGF/Sema expression patterns have and how any pro-angiogenic alterations can be therapeutically reversed.

The overlap of VEGF/Sema-based clusters with some subtypes derived from transcriptomic data suggests that existing subtypes may serve as useful biomarkers. Some evidence exists for the

performance of transcriptomic subtypes as anti-VEGF biomarkers. Bevacizumab given in the neoadjuvant setting was more effective in triple-negative breast cancers than in hormone receptor-positive breast cancers [106].

4 Dysregulation of the VEGF and Semaphorin Ligand-Receptor Families in Prostate Cancer Metastasis

4.1 Summary

The vascular endothelial growth factor (VEGF) family is central to cancer angiogenesis. However, targeting VEGF as an anti-cancer therapeutic approach has shown success for some tumor types but not others. Here we examine the expression of the expanded VEGF family in prostate cancer, including the Semaphorin (Sema) family members that compete with VEGFs for Neuropilin binding and can themselves have pro- or anti-angiogenic activity. First, we used multivariate statistical methods, including partial least squares and clustering, to examine VEGF/Sema gene expression variability in previously published prostate cancer microarray datasets. We show that unlike some cancers, such as kidney cancer, primary prostate cancer is characterized by both a down-regulation of the pro-angiogenic members of the VEGF family and a down-regulation of anti-angiogenic members of the Sema family. We found pro-lymphangiogenic signatures, including the genes encoding VEGFC and VEGFD, associated with primary tumors that ultimately became aggressive. In contrast to primary prostate tumors, prostate cancer metastases showed increased expression of key pro-angiogenic VEGF family members and further repression of anti-angiogenic class III Sema family members. Given the lack of success of VEGF-targeting molecules so far in prostate cancer, this suggests that the reduction in anti-angiogenic Sema signaling may potentiate VEGF signaling and even promote resistance to VEGF-targeting therapies. Inhibition of the VEGF 'accelerator' may need to be accompanied by promotion of the Sema 'brake' to block cancer angiogenesis. To leverage our mechanistic understanding, and to link multigene expression changes to outcomes, we performed individualized computational simulations of competitive VEGF and Sema receptor binding across

many tumor samples. The simulations suggest that loss of Sema expression promotes angiogenesis by lowering plexin signaling, not by potentiating VEGF signaling via relaxation of competition. The model also predicts that targeting Neuropilins might be problematic due to reduction in anti-angiogenesis signaling of Plexins confounding the reduction in pro-angiogenesis signaling of VEGFRs.

4.2 Methods

A flowchart of the overall procedure of this study is shown in **Figure 4-1**.

4.2.1 Data

The datasets used in this study (**Table 4-1**), were obtained from the TCGA website (http://cancergenome.nih.gov/) or from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). The majority of samples were from primary, untreated tumors, but some datasets also included normal prostate tissue samples and tumors that had metastasized from the prostate to various locations in the body. The data for these samples were included in the analysis, in separate groups distinct from the primary tumors. Outcome data (time until death) was not available, with the exception of some TCGA samples. The TCGA and GSE21034 datasets included biochemical recurrence (BCR) data, with times indicating the duration between sample collection and either BCR or BCR-free follow-up.

4.2.2 Comparison of prostate tissue types

To assess univariate differences in gene expression across prostate tissue types (normal tissue, primary tumors, and metastatic tumors), we performed Welch two-sample unpaired *t*-tests in R using the *t.test* function. P-values were subjected to multiple testing correction using the Benjamini-Hochberg procedure [117].

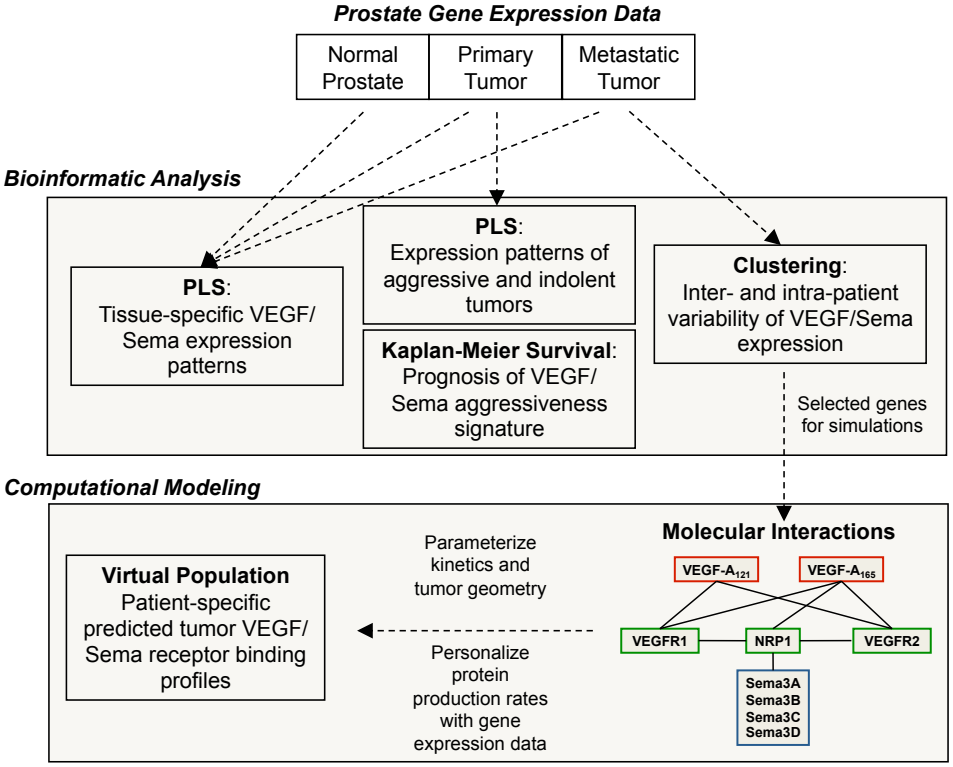

Figure 4-1: Flowchart of methods used in prostate cancer analysis.

All three tissue types were used for PLS-DA of tissue-specific differences in VEGF/Sema expression. Only primary tumor data was relevant to predicting time to biochemical recurrence.

Table 4-1: Prostate cancer datasets.

Dataset	Platform	Tissue Types	Notes	Reference
TCGA	RNA-seq	44 normal 176 primary	Copy number, miRNA, BCR	cancergenome.nih.gov
GSE21034	Exon 1.0 Array	29 normal 131 primary 19 metastasis	Copy number and BCR	[118]
GSE6919	HG U95Av2	18 donor normal 63 adjacent normal 65 primary 25 metastasis	Some adjacent normal and primary tumors are paired	[119]
GSE32269	HG U133A	22 primary 29 bone metastasis	Metastases from bone marrow biopsies	[120]
GSE38241	Agilent Whole Human Genome 4x44K	21 normal donors 18 metastases from 5 patients	Rapid autopsies; multiple metastases per patient	[121]
GSE35988	Agilent Whole Human Genome 4x44K	12 normal 49 primary 27 metastasis	Rapid autopsies	[122]

To compare prostate cancer tissue types (normal, primary tumor, metastasis), we used PLS-DA, which finds latent variables that are aligned with the direction of most co-variability between the gene expression data and the output class variable. We used the *plsda* function in the *mixOmics* package in R. The function results in a decomposition of the gene expression matrix $\mathbf{X} = \mathbf{TP} + \mathbf{E}$, where $\mathbf{T}$ is the scores matrix, $\mathbf{P}$ is the loadings matrix, and $\mathbf{E}$ is a residuals matrix. The function also decomposes the output matrix $\mathbf{Y} = \mathbf{TQ} + \mathbf{F}$, where $\mathbf{Q}$ is the Y-loadings matrix and $\mathbf{F}$ is a residuals matrix. A PLS weights matrix, $\mathbf{W}$, is also returned for predicting the scores from the gene expression data, $\mathbf{T} = \mathbf{XW}$. Thus, if an output is unknown for a particular sample, the output can be predicted from the gene expression data using $\mathbf{Y} = \mathbf{XWQ'}$.

Training errors were determined by training a PLS-DA classifier using all samples, and then performing classification on the samples. The number of misclassified samples divided by the total number of samples was the training error. To estimate the generalization error of the PLS-DA classifiers, we used leave-one-out cross-validation (LOOCV). The generalization error was the number of misclassified samples divided by the total number of samples. The PLS-DA classification was performed n times, where n was the number of samples in the data matrix. At each iteration, one sample was held out and a classifier was trained on the data consisting of n-1 samples. Then, the expression data for the held out sample was used to predict the class of the held out sample using **Y** = **XWQ'**.

We used three types of plots to show differences in gene expression between tissue types (e.g. **Figures 4-3A** through **4-3D**): density plots show differences in the shapes of the distributions, with the densities estimated using the R function *density* with default kernel bandwidth; box plots show the range of variation with statistics such as the median and quartiles; and spike plots showed the individual points to emphasize the range.

4.2.3 Prognostic significance of PLS-DA biomarkers

Data for time to follow-up or biochemical recurrence (BCR) was used to analyze survival of patients in distinct clinical or PLS-DA-derived groups. Log rank tests were used to determine differences in the Kaplan-Meier survival estimates between two classes, and Cox proportional hazard models were used to compare the effects of multiple continuous variables. We used the *survfit* and *coxph* functions in the *survival* package in R.

4.2.4 Comparison of multiple tumor samples from single patient

To assess the similarity of tumors from different metastatic sites within the same patient, we scaled the metastatic tumor data in GSE38241 by subtracting the mean expression of each gene in the normal samples. Then we grouped the metastasis samples with *K*-means clustering using the *kmeans* function in R. The *kmeans* function was executed 20 times to avoid local

minima; the cluster membership that gave the lowest within-cluster sum of squares was the one that the *kmeans* function returned. The number of clusters to be found is pre-specified. To choose the appropriate number of clusters, consensus clustering is performed. This entails clustering 100 times on random subsets of the data. This process results in a consensus matrix, which describes the co-clustering frequency of any two samples across consensus runs. Ideally, all values in this matrix will either be 1 (always in the same cluster) or 0 (never in the same cluster). The appropriate number of clusters is chosen as the value of K that has the most consensus matrix values near 1 and 0.

4.2.5 Other gene expression-based biomarkers

We also analyzed other biomarkers using the PLS-DA methodology just described. Some were based on commercially available prostate cancer diagnostics tests, such as OncotypeDx, Decipher, and Prolaris. In these cases, we used the genes in the diagnostic test in PLS-DA models, rather than the proprietary algorithms.

4.2.6 Simulation of VEGF/Sema binding

To assess the effects of patient tumor gene expression on VEGF and semaphorin signaling, we built a model of ligand-receptor binding that encompassed all VEGF proteins for which gene expression data was available, all class 3 semaphorins, and all receptors known to bind these ligands. We expanded our validated kinetic model comprising interactions between two isoforms of VEGF-A ($VEGF_{121}$ and $VEGF_{165}$), VEGFR-1, VEGFR-2, Neuropilin-1, Neuropilin-2 and soluble VEGFR-1 (sVEGFR-1) [47,113,123,124] to include the additional encoded VEGF and Sema family proteins. The scope of some of these modeling studies has included the entire body so that drug pharmacokinetics could be studied [47,123]; here, we limit the scope of the model to a single compartment representing a tumor and the surrounding microvasculature, justified by our results showing the small size of interaction effects between the

tumor compartment and other compartments. We use this model to study only the effects of gene expression variation.

Geometry. We assumed that the tumor volume consisted only of tumor cells, endothelial cells, and the interstitial space between these cells. The interstitial space available to ligands and the surface area to volume ratios of the two cell types were as previously described ([113] and **Table 4-2**).

Table 4-2: Geometric parameters.

Parameter	Value	Units	Ref.
Microvessel surface area to volume ratio	105	cm^2/cm^3 tissue	[113]
Tumor surface area to volume ratio	1534	cm^2/cm^3 tissue	[113]
Endothelial cell surface area	$1 \bullet 10^{-5}$	cm^2	[113]
Tumor cell surface area	$1 \bullet 10^{-5}$	cm^2	[113]
Interstitial volume fraction	0.58	cm^3/cm^3 tissue	[113]

Kinetics. Rate constants for the association and dissociation reactions between the two major VEGF-A isoforms, VEGF-A_{165} and VEGF-A_{121}, and VEGFR-1, VEGFR-2, and Neuropilin-1 were as previously described ([124], **Tables 4-3** and **4-4**). The interaction of VEGF-A_{165} with Neuropilin-2 and coupling between Neuropilin-2 and VEGFR-1/VEGFR-2 were as described in [47] (**Tables 4-3** and **4-4**). Interactions of sVEGFR-1 with the two VEGF-A isoforms and Neuropilin-1 were as described in [123] (**Table 4-3**).

Table 4-3: Rate constants of ligand-receptor binding reactions.

Reaction	k_{on} ($M^{-1}sec^{-1}$)	k_{off} (sec^{-1})	K_d (pM)	Ref
$VEGFA_{165}$ + VEGFR1 ↔ $VEGFA_{165}$-R1	$3•10^7$	$1•10^{-3}$	33	[113]
$VEGFA_{165}$ + VEGFR2 ↔ $VEGFA_{165}$-R2	$1•10^7$	$1•10^{-3}$	100	[113]
$VEGFA_{165}$ + NRP1 ↔ $VEGFA_{165}$-NRP1	$3.125•10^6$	$1•10^{-3}$	320	[113]
$VEGFA_{165}$ + NRP2 ↔ $VEGFA_{165}$-NRP2	$1•10^6$	$1•10^{-3}$	320	[113]
$VEGFA_{121}$ + VEGFR1 ↔ $VEGFA_{121}$-R1	$3•10^7$	$1•10^{-3}$	33	[113]
$VEGFA_{121}$ + VEGFR2 ↔ $VEGFA_{121}$-R2	$1•10^7$	$1•10^{-3}$	100	[113]
$VEGFA_{121}$ + R1-NRP1 ↔ $VEGFA_{121}$-R1-NRP1	$3•10^7$	$1•10^{-3}$	33	[113]
$VEGFA_{121}$ + R1-NRP2 ↔ $VEGFA_{121}$-R1-NRP2	$3•10^7$	$1•10^{-3}$	33	[113]
PlGF1 + VEGFR1 ↔ PlGF1-R1	$4.3•10^6$	$1•10^{-3}$	230	[125]
PlGF1 + R1-NRP1 ↔ PlGF1-R1-NRP1	$4.3•10^6$	$1•10^{-3}$	230	[b]
PlGF1 + R1-NRP2 ↔ PlGF1-R1-NRP2	$4.3•10^6$	$1•10^{-3}$	230	[b]
PlGF2 + VEGFR1 ↔ PlGF2-R1	$4.3•10^6$	$1•10^{-3}$	230	[125]
PlGF2 + NRP1 ↔ PlGF2-NRP1	$3.125•10^6$	$1•10^{-3}$	320	[a]
PlGF2 + NRP2 ↔ PlGF2-NRP2	$3.125•10^6$	$1•10^{-3}$	320	[a]
$VEGFB_{167}$ + VEGFR1 ↔ $VEGFB_{167}$-R1	$4.3•10^6$	$1•10^{-3}$	230	[b]
$VEGFB_{186}$ + VEGFR1 ↔ $VEGFB_{186}$-R1	$4.3•10^6$	$1•10^{-3}$	230	[b]
$VEGFB_{167}$ + NRP1 ↔ $VEGFB_{167}$-NRP1	$3.125•10^6$	$1•10^{-3}$	320	[a]
$VEGFB_{186}$ + NRP1 ↔ $VEGFB_{186}$-NRP1	$3.125•10^6$	$1•10^{-3}$	320	[a]
VEGFC + VEGFR2 ↔ VEGFC-R2	$2.4•10^6$	$1•10^{-3}$	410	[126]
VEGFC + VEGFR3 ↔ VEGFC-R3	$7.4•10^6$	$1•10^{-3}$	135	[126]
VEGFC + NRP2 ↔ VEGFC-NRP2	$3.125•10^6$	$1•10^{-3}$	320	[a]
VEGFD + VEGFR2 ↔ VEGFD-R2	$2.4•10^6$	$1•10^{-3}$	410	[c]
VEGFD + VEGFR3 ↔ VEGFD-R3	$7.4•10^6$	$1•10^{-3}$	135	[c]
VEGFD + NRP2 ↔ VEGFD-NRP2	$3.125•10^6$	$1•10^{-3}$	320	[a]
Sema3A + NRP1 ↔ Sema3A-NRP1	$1•10^6$	$1•10^{-3}$	1000	[127]
Sema3B + NRP1 ↔ Sema3B-NRP1	$1•10^6$	$1•10^{-3}$	1000	[d]
Sema3B + NRP2 ↔ Sema3B-NRP2	$1•10^6$	$1•10^{-3}$	1000	[d]
Sema3C + NRP1 ↔ Sema3C-NRP1	$7.7•10^5$	$1•10^{-3}$	1300	[127]
Sema3C + NRP2 ↔ Sema3C-NRP2	$5.9•10^5$	$1•10^{-3}$	1700	[127]
Sema3D + NRP1 ↔ Sema3D-NRP1	$1•10^6$	$1•10^{-3}$	1000	[d]
Sema3D + NRP2 ↔ Sema3D-NRP2	$1•10^6$	$1•10^{-3}$	1000	[d]
Sema3E + PLXND1 ↔ Sema3E-PD1	$7.7•10^6$	$1•10^{-3}$	130	[86]
Sema3F + NRP2 ↔ Sema3F-NRP2	$6.7•10^5$	$1•10^{-3}$	1500	[127]
Sema3G + NRP2 ↔ Sema3G-NRP2	$1•10^6$	$1•10^{-3}$	1000	[d]
Sema3A + PlxnAi-NRP1 ↔ Sema3A-NRP1-PAi	$6.3•10^6$	$1•10^{-3}$	190	[127][e]
Sema3A + PlxnD1-NRP1 ↔ Sema3A-NRP1-PD1	$1•10^6$	$1•10^{-3}$	1000	[d]

Table 4-3 continued on next page

Sema3B + PlxnAi-NRP1 ↔ Sema3B-NRP1-PAi	$1 \bullet 10^{6}$	$1 \bullet 10^{-3}$	1000	[d]
Sema3B + PlxnAi-NRP2 ↔ Sema3B-NRP2-PAi	$1 \bullet 10^{6}$	$1 \bullet 10^{-3}$	1000	[d]
Sema3C + PlxnAi-NRP1 ↔ Sema3C-NRP1-PAi	$8.5 \bullet 10^{5}$	$1 \bullet 10^{-3}$	1200	[127] [e]
Sema3C + PlxnAi-NRP2 ↔ Sema3C-NRP2-PAi	$3.9 \bullet 10^{5}$	$1 \bullet 10^{-3}$	2300	[127] [e]
Sema3C + PlxnD1-NRP1 ↔ Sema3C-NRP1-PD1	$1 \bullet 10^{6}$	$1 \bullet 10^{-3}$	1000	[d]
Sema3C + PlxnD1-NRP2 ↔ Sema3C-NRP2-PD1	$1 \bullet 10^{6}$	$1 \bullet 10^{-3}$	1000	[d]
Sema3D + PlxnAi-NRP1 ↔ Sema3D-NRP1-PAi	$1 \bullet 10^{6}$	$1 \bullet 10^{-3}$	1000	[d]
Sema3D + PlxnAi-NRP2 ↔ Sema3D-NRP2-PAi	$1 \bullet 10^{6}$	$1 \bullet 10^{-3}$	1000	[d]
Sema3F + PlxnAi-NRP2 ↔ Sema3F-NRP2-PAi	$2.7 \bullet 10^{6}$	$1 \bullet 10^{-3}$	440	[127] [e]
Sema3G + PlxnAi-NRP2 ↔ Sema3G-NRP2-PAi	$1 \bullet 10^{6}$	$1 \bullet 10^{-3}$	1000	[d]
$VEGF_{165}$ + sVEGFR1 ↔ $VEGF_{165}$-sR1	$3 \bullet 10^{7}$	$1 \bullet 10^{-3}$	33	[123]
$VEGF_{121}$ + sVEGFR1 ↔ $VEGF_{121}$-sR1	$3 \bullet 10^{7}$	$1 \bullet 10^{-3}$	33	[123]
PlGF1 + sVEGFR1 ↔ PlGF1-sR1	$4.3 \bullet 10^{6}$	$1 \bullet 10^{-3}$	230	[b]
PlGF2 + sVEGFR1 ↔ PlGF2-sR1	$4.3 \bullet 10^{6}$	$1 \bullet 10^{-3}$	230	[b]
$VEGFB_{167}$ + sVEGFR1 ↔ $VEGFB_{167}$-sR1	$4.3 \bullet 10^{6}$	$1 \bullet 10^{-3}$	230	[b]
$VEGFB_{186}$ + sVEGFR1 ↔ $VEGFB_{186}$-sR1	$4.3 \bullet 10^{6}$	$1 \bullet 10^{-3}$	230	[b]
sVEGFR1 + NRP1 ↔ sR1-NRP1	$5.56 \bullet 10^{6}$	$1 \bullet 10^{-2}$	1800	[123]
sVEGFR1 + NRP2 ↔ sR1-NRP2	$5.56 \bullet 10^{6}$	$1 \bullet 10^{-2}$	1800	[f]
$VEGF_{121}$ + sR1-NRP1 ↔ $VEGF_{121}$-sR1-NRP1	$3 \bullet 10^{7}$	$1 \bullet 10^{-3}$	33	[123]
PlGF1 + sR1-NRP1 ↔ PlGF1-sR1-NRP1	$4.3 \bullet 10^{6}$	$1 \bullet 10^{-3}$	230	[b]
$VEGF_{121}$ + sR1-NRP2 ↔ $VEGF_{121}$-sR1-NRP2	$3 \bullet 10^{7}$	$1 \bullet 10^{-3}$	33	[123]
PlGF1 + sR1-NRP2 ↔ PlGF1-sR1-NRP2	$4.3 \bullet 10^{6}$	$1 \bullet 10^{-3}$	230	[b]
$VEGF_{121}$-sR1 + NRP1 ↔ $VEGF_{121}$-sR1-NRP1	$5.56 \bullet 10^{6}$	$1 \bullet 10^{-2}$	1800	[123]
PlGF1-sR1 + NRP1 ↔ PlGF1-sR1-NRP1	$5.56 \bullet 10^{6}$	$1 \bullet 10^{-2}$	1800	[f]
$VEGF_{121}$-sR1 + NRP2 ↔ $VEGF_{121}$-sR1-NRP2	$5.56 \bullet 10^{6}$	$1 \bullet 10^{-2}$	1800	[f]
PlGF1-sR1 + NRP2 ↔ PlGF1-sR1-NRP2	$5.56 \bullet 10^{6}$	$1 \bullet 10^{-2}$	1800	[f]
$VEGF_{165}$ + ECM ↔ $VEGF_{165}$-ECM	$4.2 \bullet 10^{5}$	$1 \bullet 10^{-2}$	23800	[113]
PlGF2 + ECM ↔ PlGF2-ECM	$1 \bullet 10^{7}$	$1 \bullet 10^{-2}$	1000	[128][g]
sVEGFR1 + ECM ↔ sR1-ECM	$4.2 \bullet 10^{5}$	$1 \bullet 10^{-2}$	23800	[123]

[a] No K_d data available; assumed same as K_d for $VEGFA_{165}$-NRP1
[b] No K_d data available; assumed same as K_d for PlGF-VEGFR1
[c] No K_d data available; assumed same as K_ds for VEGFC binding to VEGFR-2 and VEGFR-3
[d] No K_d data available; assumed K_d of 1 nM (the middle of the range of available Sema3 K_ds)
[e] The K_d listed is an overall K_d for the cells expressing both NRP1 and PlxnA1
[f] Assumed to be the same as sVEGFR1 binding to NRP1
[g] In [128], PlGF2 on average has 26-fold lower affinity for ECM proteins that $VEGFA_{165}$

Table 4-4: Rate constants of receptor coupling reactions.

Reaction	k_c $((mol/cm^2)^{-1}sec^{-1})$	k_{dis} (sec^{-1})	Reference
VEGFR1 + NRP1 ↔ R1-NRP1	$1•10^{14}$	$1•10^{-2}$	[113]
VEGFR1 + NRP2 ↔ R1-NRP2	$1•10^{14}$	$1•10^{-2}$	[47]
PlxnAi + NRPj ↔ PAi-NRPj	$1•10^{14}$	$1•10^{-2}$	[a]
PlxnD1 + NRPj ↔ PD1-NRPj	$1•10^{14}$	$1•10^{-2}$	[a]
$VEGF_{121}$-R1 + NRP1 ↔ $VEGF_{121}$-R1-NRP1	$1•10^{14}$	$1•10^{-2}$	[113]
PlGF1-R1 + NRP1 ↔ PlGF1-R1-NRP1	$1•10^{14}$	$1•10^{-2}$	[113]
$VEGF_{165}$-R2 + NRP1 ↔ $VEGF_{165}$-R2-NRP1	$3.1•10^{13}$	$1•10^{-3}$	[113]
$VEGF_{165}$-NRP1 + VEGFR2 ↔ $VEGF_{165}$-R2-NRP1	$1•10^{14}$	$1•10^{-3}$	[113]
$VEGF_{165}$-R2 + NRP2 ↔ $VEGF_{165}$-R2-NRP2	$3.1•10^{13}$	$1•10^{-3}$	[47]
$VEGF_{165}$-NRP2 + VEGFR2 ↔ $VEGF_{165}$-R2-NRP2	$1•10^{14}$	$1•10^{-3}$	[47]
VEGFC-R2 + NRP2 ↔ VEGFC-R2-NRP2	$3.1•10^{13}$	$1•10^{-3}$	[47]
VEGFC-NRP2 + VEGFR2 ↔ VEGFC-R2-NRP2	$1•10^{14}$	$1•10^{-3}$	[47]
VEGFC-R3 + NRP2 ↔ VEGFC-R3-NRP2	$3.1•10^{13}$	$1•10^{-3}$	[47]
VEGFC-NRP2 + VEGFR3 ↔ VEGFC-R3-NRP2	$1•10^{14}$	$1•10^{-3}$	[47]
VEGFD-R2 + NRP2 ↔ VEGFD-R2-NRP2	$3.1•10^{13}$	$1•10^{-3}$	[47]
VEGFD-NRP2 + VEGFR2 ↔ VEGFD-R2-NRP2	$1•10^{14}$	$1•10^{-3}$	[47]
VEGFD-R3 + NRP2 ↔ VEGFD-R3-NRP2	$3.1•10^{13}$	$1•10^{-3}$	[47]
VEGFD-NRP2 + VEGFR3 ↔ VEGFD-R3-NRP2	$1•10^{14}$	$1•10^{-3}$	[47]
Sema3A-NRP1 + PlxnAi ↔ Sema3A-NRP1-PAi	$1•10^{14}$	$1•10^{-2}$	[a]
Sema3A-NRP1 + PlxnD1 ↔ Sema3A-NRP1-PD1	$1•10^{14}$	$1•10^{-2}$	[a]
Sema3B-NRP1 + PlxnAi ↔ Sema3B-NRP1-PAi	$1•10^{14}$	$1•10^{-2}$	[a]
Sema3B-NRP2 + PlxnAi ↔ Sema3B-NRP2-PAi	$1•10^{14}$	$1•10^{-2}$	[a]
Sema3C-NRP1 + PlxnAi ↔ Sema3C-NRP1-PAi	$1•10^{14}$	$1•10^{-2}$	[a]
Sema3C-NRP1 + PlxnD1 ↔ Sema3C-NRP1-PD1	$1•10^{14}$	$1•10^{-2}$	[a]
Sema3C-NRP2 + PlxnAi ↔ Sema3C-NRP2-PAi	$1•10^{14}$	$1•10^{-2}$	[a]
Sema3C-NRP2 + PlxnD1 ↔ Sema3C-NRP2-PD1	$1•10^{14}$	$1•10^{-2}$	[a]
Sema3D-NRP1 + PlxnAi ↔ Sema3D-NRP1-PAi	$1•10^{14}$	$1•10^{-2}$	[a]
Sema3D-NRP2 + PlxnAi ↔ Sema3D-NRP2-PAi	$1•10^{14}$	$1•10^{-2}$	[a]
Sema3F-NRP2 + PlxnAi ↔ Sema3F-NRP2-PAi	$1•10^{14}$	$1•10^{-2}$	[a]

Table 4-4 continued on next page

Sema3G-NRP2 + PlxnAi ↔ Sema3G-NRP2-PAi	$1 \bullet 10^{14}$	$1 \bullet 10^{-2}$	[a]
The letter i refers to the four class A Plexins The letter j refers to the two Neuropilins [a] Assumed diffusion-limited kinetics when no experimental data was available			

Table 4-5: Target free ligand concentrations.

Protein	Molecular Weight (kDa)	Target Concentration (pg/mL)	Target Concentration (pM)	Reference
VEGF-A_{165}	45	100	2.22	[9,10,11,13]
VEGF-A_{121}	45	100	2.22	[9,10,11,13]
sVEGFR1	110	100	0.91	[129,130,131]
PlGF-1	56	10	0.18	[125,132,133]
PlGF-2	64	10	0.16	[125,132,133]
VEGF-B_{167}	42	100	2.4	[134]
VEGF-B_{186}	60	100	1.7	[134]
VEGF-C	26	1000	38	[132]
VEGF-D	26	100	3.8	[135]
Sema3A	86	10000	116	[a]
Sema3B	81	10000	123	[b]
Sema3C	83	10000	120	[b]
Sema3D	86	10000	116	[b]
Sema3E	86	10000	116	[b]
Sema3F	86	10000	116	[b]
Sema3G	84	10000	119	[b]
[a] No cancer-specific measurements available, but [136] has healthy measurements [b] Assumed the same order of magnitude as Sema3A				

Table 4-6: Model parameters.

Parameter	Value	Units	Reference
VEGF-A_{165} secretion rate [a]	0.027	#/cell/sec	[b]
VEGF-A_{121} secretion rate [a]	0.013	#/cell/sec	[b]
sVEGFR1 secretion rate [a]	0.059	#/cell/sec	[b]
PlGF-1 secretion rate [a]	0.00026	#/cell/sec	[b]
PlGF-2 secretion rate [a]	0.00011	#/cell/sec	[b]
VEGF-B_{167} secretion rate [a]	0.017	#/cell/sec	[b]
VEGF-B_{186} secretion rate [a]	0.029	#/cell/sec	[b]
VEGF-C secretion rate [a]	0.96	#/cell/sec	[b]
VEGF-D secretion rate [a]	0.096	#/cell/sec	[b]
Sema3A secretion rate [a]	0.97	#/cell/sec	[b]
Sema3B secretion rate [a]	1.50	#/cell/sec	[b]
Sema3C secretion rate [a]	1.02	#/cell/sec	[b]
Sema3D secretion rate [a]	1.41	#/cell/sec	[b]
Sema3E secretion rate [a]	0.0046	#/cell/sec	[b]
Sema3F secretion rate [a]	0.52	#/cell/sec	[b]
Sema3G secretion rate [a]	0.69	#/cell/sec	[b]
Endothelial VEGFR1	3,750	#/cell	[47]
Endothelial VEGFR2	300	#/cell	[47]
Endothelial VEGFR3	1,000	#/cell	[c]
Endothelial NRP1	20,000	#/cell	[47]
Endothelial NRP2	20,000	#/cell	[47]
Endothelial PLXNA1	1,000	#/cell	[c]
Endothelial PLXNA2	1,000	#/cell	[c]
Endothelial PLXNA3	1,000	#/cell	[c]
Endothelial PLXNA4	1,000	#/cell	[c]
Endothelial PLXND1	1,000	#/cell	[c]
Tumor VEGFR1 [a]	1,100	#/cell	[47]
Tumor VEGFR2 [a]	550	#/cell	[47]
Tumor VEGFR3 [a]	1,000	#/cell	[c]
Tumor NRP1 [a]	39,500	#/cell	[47]
Tumor NRP2 [a]	39,500	#/cell	[47]
Tumor PLXNA1[a]	1,000	#/cell	[c]
Tumor PLXNA2[a]	1,000	#/cell	[c]
Tumor PLXNA3[a]	1,000	#/cell	[c]
Tumor PLXNA4[a]	1,000	#/cell	[c]
Tumor PLXND1[a]	1,000	#/cell	[c]
ECM concentration	0.75	μM	[113]
Receptor internalization rate	$2.8 \bullet 10^{-4}$	sec^{-1}	[113]

[a] These parameters vary based on expression of the corresponding genes.
[b] These parameters were tuned to yield the target concentrations in Table S11.
[C] No experimental data for Plexin cell surface densities are available, therefore 1,000 was chosen as it falls in the middle of the range of receptor densities.

For reactions whose rate constants were not directly known, we assumed a k_{off} of $1x10^{-3}$ sec^{-1} (consistent with measured off rates for this family) and calculated k_{on} as k_{off}/K_d, where K_d is the measured dissociation constant obtained from the literature. Dissociation constants for the interactions of PlGF-1 and PlGF-2 with VEGFR-1 were available [125], but no data was available for the binding of PlGF-2 to Neuropilin-1, thus it was estimated to be the same as for the binding of VEGF-A_{165} to Neuropilin-1. Dissociation constants were not available for the two isoforms of VEGF-B, VEGF-B_{167} and VEGF-B_{186}, for their receptors VEGFR-1 and Neuropilin-1. We used the same rate constants for these two reactions as PlGF-2. Dissociation constants were available for the binding of VEGF-C to VEGFR-2 and VEGFR-3 [126], but not for the binding of VEGF-C to Neuropilin-2. In this case, we assumed the same kinetics as the binding of VEGF-A_{165} to Neuropilin-1. No dissociation constants were available for the binding of VEGF-D to its receptors, VEGFR-2, VEGFR-3, and Neuropilin-2, therefore we used the same rate constants as VEGF-C.

For class 3 Semaphorin binding to Neuropilins and Neuropilin/Plexin complexes, dissociation constant data was only available for the binding of Sema3A and Sema3C to Neuropilin-1 and Neuropilin-1/Plexin-A1 complexes, and for the binding of Sema3C and Sema3F to Neuropilin-2 and Neuropilin-2/Plexin-A1 complexes ([127] and **Table 4-3**). The on rate constants k_{on} for the binding of these Semaphorins to their respective Neuropilins could be estimated directly from the dissociation constants, but the values of k_{on} for binding of Semaphorins to Neuropilin/Plexin complexes had to be determined from simulations due to the presence of both Neuropilin and Neuropilin/Plexin complexes. We assumed that the kinetics of coupling between the Neuropilin and Plexin receptors occurred with identical rates constants to those for VEGFR1 and Neuropilin-1 coupling (**Table 4-4**). Data was also available for the binding of Sema3E to its receptor Pleixn-D1 ([86] and **Table 4-3**).

Receptor densities. As the amount of receptors on the surface of cells *in vivo* is difficult to measure, we use quantitative flow cytometry data of cultured cell lines where available. The densities of VEGFR-1, VEGFR-2, and NRP-1 have previously been measured in this way and used in computational models ([47] and **Table 4-6**). We assumed NRP-2 densities were similar to NRP-1 as in [47]. To our knowledge, VEGFR-3 and Plexin levels have not been measured on endothelial or tumor cells; therefore, we set the density of each of these receptors to 1,000, which falls in the middle of the range of cell surface densities in our model.

Ligand secretion rates. We adjusted ligand secretion rates so that the steady-state concentrations of free ligands matched pre-determined target concentrations. The target concentrations were derived from published measurements of plasma concentrations in cancer patients. Actual interstitial ligand concentrations would likely be higher than plasma concentrations, but published data of interstitial concentrations are rare due to the difficulty in obtaining these measurements. Data for plasma concentrations of VEGF ligands were widely available due to studies testing their performance as pharmacodynamic markers of VEGF-targeting drugs (**Table 4-5**). To our knowledge, only one study has been published with data for plasma Semaphorin concentrations, which found a plasma Sema3A concentration of 74.41 ng/mL in healthy patients [136]. Here, we used this order of magnitude (10^4 pg/mL) for each of the seven class 3 Semaphorins (**Table 4-5**).

Variation due to gene expression. Gene expression data was used to vary the protein production rates in tumor cells only. The log2-transformed gene expression data was first median-centered and then added to the log2-transformed nominal production rates found in the previous section. The actual production rate was then found by exponentiation with base 2. Thus the median gene expression corresponds to the nominal secretion rate found in the previous section. For *VEGFA*, *PGF*, *VEGFB*, and *FLT1*, two isoforms were represented in the model for each gene. We used TCGA data to determine average isoform fractions of total gene expression (**Figure 4-2**). Isoforms from identical genes were typically highly correlated.

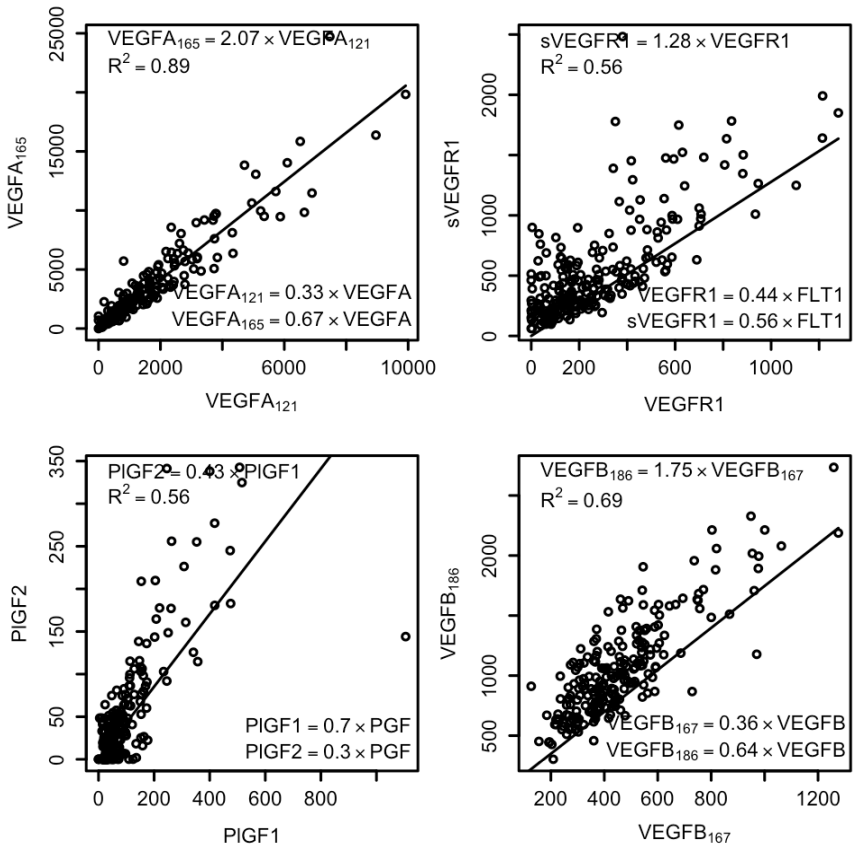

Figure 4-2: Isoform ratios of genes with alternative splicing.

Four genes known to have multiple isoforms due to alternative transcript splicing were analyzed in the TCGA RNA-Seq dataset. Linear regression models showed a high degree of correlation between pairs of isoforms originating from the same gene. From the linear models, we calculated the relative fraction of total gene expression accounted for by each isoform (bottom right of each plot). These values were used to determine isoform secretion rates in the mathematical model of tumor VEGF/Semaphorin receptor binding.

The system of coupled nonlinear differential equations describing the amounts of each molecule/complex was solved numerically in Fortran using the fifth-order Runge-Kutta method with adaptive step-size control.

4.3 Results

4.3.1 Primary prostate tumor VEGF/Sema alterations

To examine why prostate cancer is less susceptible to VEGF inhibition, we first compared gene expression between primary prostate tumors and a cancer type that typically does respond to VEGF inhibitors, renal cell carcinoma [137]. Using the TCGA RNA-Seq dataset, we found that while two key pro-angiogenic ligands, *VEGFA* and *PGF*, were up-regulated in renal cell carcinoma, they were down-regulated in primary prostate adenocarcinoma (**Figure 4-3A through 4-3D**). This could indicate a lack of VEGF signaling for VEGF inhibitors to target, making attempts at targeting the VEGF pathway in prostate cancer futile. This pattern was observed for several other VEGF ligands and receptors: these genes were down-regulated or unchanged in prostate cancer whereas they were up-regulated in renal cell carcinoma (**Figure 4-3E** for prostate cancer; renal cell carcinoma not shown). However, the up- or down-regulation of VEGF ligands and receptors in cancer types may not fully explain the response to VEGF inhibitors in prostate cancer. Subsets of tumors may have pro-angiogenic gene expression despite the anti-angiogenic pattern in the tumors as an overall group. Additionally, other pathways that interact with the VEGF pathway may affect signaling.

We considered an additional set of genes that have been demonstrated to affect VEGF signaling: the semaphorins and their plexin receptors, bringing the total number of VEGF/Sema-related genes under consideration to 39. We expanded our comparison of normal prostate tissue and primary prostate tumors to include multiple microarray datasets in addition to the TCGA dataset. For primary tumor versus normal tissue comparisons, we measured differences in gene expression using a two-tailed *t*-test with multiple testing correction using the Benjamini-

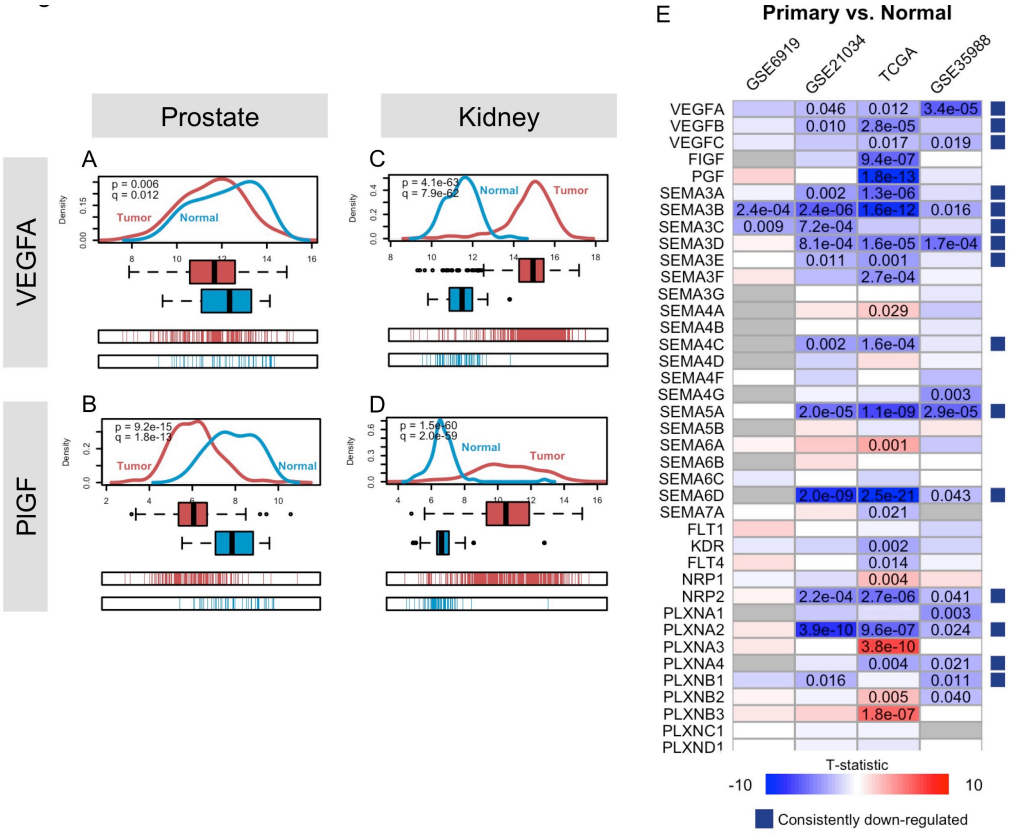

Figure 4-3: Down-regulation of pro- and anti-angiogenic ligands in primary prostate tumors.

A-B: *VEGFA* (A) and *PGF* (B) are expressed at lower levels in prostate tumors than in normal prostate tissue in the TCGA dataset, as shown by the density plots (top), box plots (middle), and spike plots (bottom). **C-D:** This contrasts with the TCGA renal cell carcinoma (kidney) dataset, where *VEGFA* (C) and *PGF* (D) are heavily up-regulated in tumors. **E,** *VEGFA* down-regulation in primary prostate cancer is observed across TCGA and microarray datasets, as is consistent down-regulation of class 3 semaphorins. The number in the boxes indicates the two-tailed *t*-test p-value after multiple testing correction with the Benjamini-Hochberg procedure. Only comparisons with corrected p-values less than 0.05 are displayed. The colors of the boxes indicate the magnitude of the *t*-statistic. The blue boxes to the right of the rows indicate that a gene is significantly down-regulated in two or more datasets with no significant differences in other datasets.

Hochberg procedure. We found that in addition to VEGF ligands, many semaphorins were also down-regulated in prostate cancer (**Figure 4-3E**). In particular, class 3 semaphorins, which have potential anti-angiogenic effects due to their ability to compete with VEGF for neuropilin binding, were down-regulated. Not all gene expression changes were consistent across all datasets; ones that were significant (defined as q-value less than 0.05) in two or more datasets were marked with blue or red boxes to the right. The subset of genes consistently down-regulated in primary tumors included three out of five VEGF ligands (*VEGFA*, *VEGFB*, and *VEGFC*) and five out of seven class 3 semaphorins (*SEMA3A*, *SEMA3B*, *SEMA3C*, *SEMA3D*, and *SEMA3E*). These results made the overall impact on angiogenesis unclear as both pro-angiogenic VEGF signals and anti-angiogenic semaphorin signals were reduced in primary tumors.

4.3.2 A pro-lymphangiogenic gene expression signature is associated with aggressive primary tumors

To assess whether some subsets of primary prostate tumors may have differing potential benefits from VEGF signaling inhibitors, we used partial least squares discriminant analysis (PLS-DA), a multivariate algorithm that allowed us to simultaneously consider both the pattern of VEGF/Sema gene expression and effects on an output variable. As an output, we used a binary variable indicating whether biochemical recurrence (BCR) eventually occurred. Data for BCR/follow-up times were available in the TCGA and GSE21034 datasets, with 10 and 27 BCR events, respectively. Since BCR-negative samples with short follow-up times may eventually undergo recurrence, we used only the 10 and 27 BCR-negative samples with the longest follow-up times in the TCGA and GSE21034 datasets, respectively. These groups of indolent samples had follow-up times greater than 3.3 years in the TCGA dataset and greater than 5.2 years in the GSE21034 dataset. The TCGA PLS-DA model was effective in differentiating aggressive and indolent tumors, with a 95% training accuracy (**Figure 4-4A**). The GSE21034 PLS-DA model, on the other hand, was less effective with only a 68.5% training accuracy (**Figure 4-4B**). This

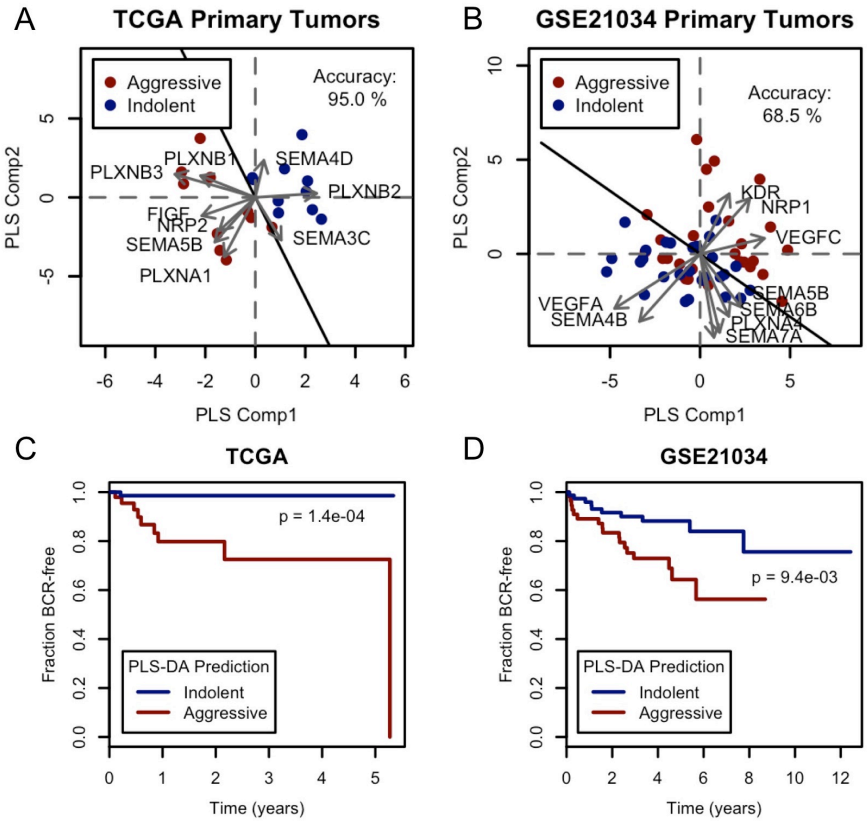

Figure 4-4: VEGF/Sema expression signatures predicting biochemical recurrence (BCR).

A-B: PLS-DA scores/loadings plots for the TCGA (A) and GSE21034 (B) datasets. The training accuracies and the discriminant line separating the two classes are displayed **C-D:** Survival curves show prognostic significance of VEGF/Sema signatures in the TCGA (C) and GSE21034 (D) datasets. The p-values are from log rank tests of the Kaplan-Meier survival estimators for each group.

discrepancy in performance between TCGA and GSE21034 datasets was seen for other biomarkers meant to distinguish aggressive from indolent tumors (data not shown). Notably, the VEGF/Sema PLS-DA model (and PLS-DA models based on the genes from other biomarkers) lost some of its predictive ability when correcting for Gleason score, but still was significantly prognostic (**Figure 4-5**). ROC curves from leave-one-out cross-validation barely deviated from a 45-degree line (data not shown), indicating that these models would likely be ineffective in distinguishing aggressive and indolent tumors in other datasets. Nonetheless, the PLS-DA models provided gene expression signatures with potential prognostic significance: when the PLS discriminant scores were used as predictive variables in Kaplan-Meier survival analysis, aggressive and indolent tumors had significantly different outcomes (**Figures 4-4C, 4-4D**).

The PLS-DA models also yielded information regarding patterns of VEGF/Sema expression associated with aggressive prostate tumors. The gene-specific loadings vectors (arrows in **Figures 4-4A, 4-4B**) showed that association of VEGF/Sema genes with aggressiveness was different between the two datasets. For the TCGA dataset, aggressive tumors were associated with high expression of *FIGF* (VEGFD), *NRP2*, *PLXNA1*, *PLXNB1*, *PLXNB3*, and *SEMA5B* and low expression of *PLXNB2* and *SEMA4D*. The first two of these genes, *FIGF* (VEGFD) and *NRP2*, mediate pro-lymphangiogenesis signals. The process of lymphangiogenesis may provide tumors with a route by which to escape their tissue of origin into the bloodstream [138]. In the GSE21034 model, high expression of *VEGFC*, *KDR,* and *NRP1* and low expression of *VEGFA* and *SEMA4B* were associated with aggressive tumors. Although this signature was different than the one found in the TCGA dataset, the high expression of *VEGFC* also suggested a potential for lymphangiogenic activity in aggressive primary tumors.

4.3.3 Metastatic prostate tumors are associated with a pro-angiogenic signature

While primary tumors are often treated with surgery and radiation, targeted therapeutics such as VEGF inhibitors are used more often in metastatic disease. Therefore, we next considered

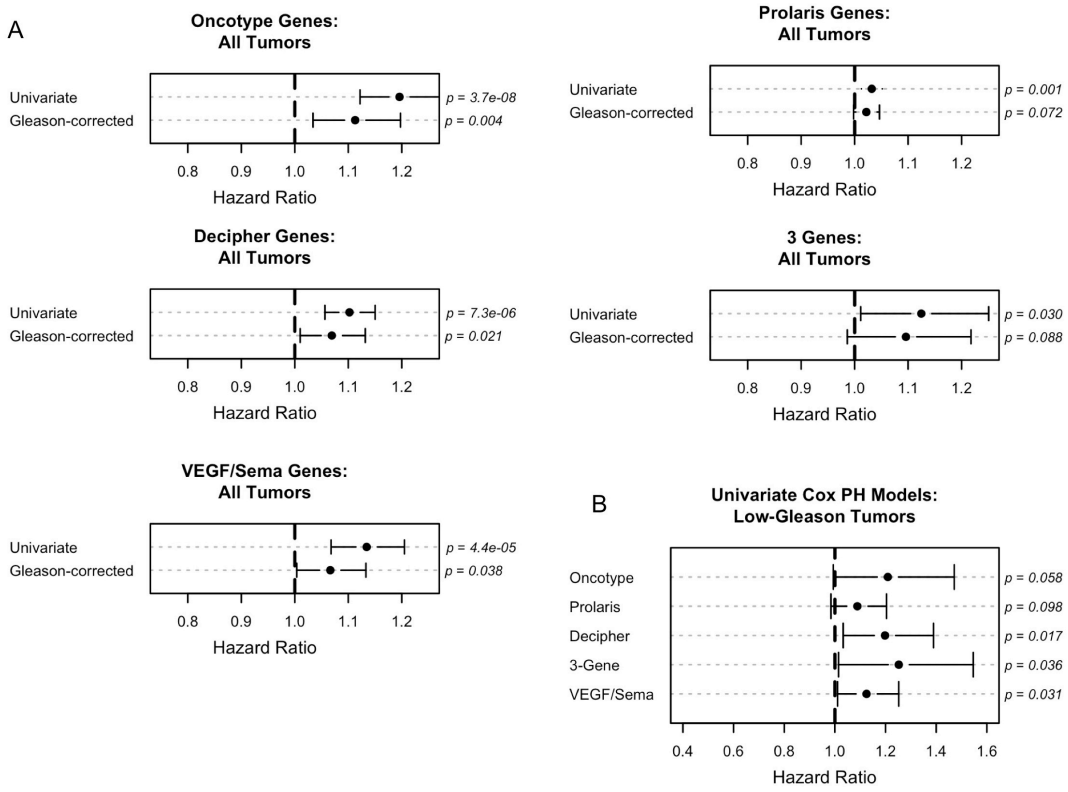

Figure 4-5: Cox proportional hazards modeling of the association between PLS-DA biomarkers and biochemical recurrence (BCR).

This figure expands upon Figure 4-4. **A:** Hazard ratios for the four biomarkers in all primary tumors of the GSE21034 dataset (n=131). Both univariate and Gleason score-corrected models are displayed. **B:** Univariate models of primary tumors with Gleason scores equal to 6 or 7 with a primary score of 3 (n=94). The four biomarkers and the VEGF/Sema PLS-DA predictor are shown.

VEGF/Sema gene expression in metastatic tumors. In contrast to the reduced expression of VEGF ligands in primary tumors, metastatic tumors tended to have higher expression of the major pro-angiogenic ligand, *VEGFA* (**Figures 4-6A, 4-6B**). This ligand was up-regulated in metastases relative to primary tumors and normal tumors in the GSE6919, GSE35988, and GSE38241 datasets, but was actually down-regulated in GSE21034 and GSE32269. Notably, metastatic samples in the three datasets with up-regulated *VEGFA* were all obtained from warm autopsy programs where samples were processed rapidly upon the death of the patient. Metastatic samples in GSE32269 were from bone marrow biopsies of live patients, and no details were given regarding how metastatic samples were obtained in the GSE21034 dataset. The class 3 semaphorins were down-regulated in metastases relative to normal prostate tissue (**Figure 4-6B**), suggesting that the loss of semaphorin expression in primary prostate tumors was maintained upon metastasis. *SEMA3C* was further down-regulated relative to primary tumors as well. Other expression alterations recurrent across datasets included up-regulation of *NRP1*, *PLXNA1*, and *PLXNA3* relative to both normal tissue and primary tumors. These three genes participate in class 3 semaphorin signaling, while only *NRP1* participates in VEGF signaling. *KDR* and *NRP2* were recurrently down-regulated in metastases relative to normal tissue but not relative to primary tumors.

We used PLS-DA to show that the VEGF/Sema gene expression alterations present in metastatic prostate tumors differed significantly from both normal tissue and primary tumors. PLS-DA models comparing metastases with primary tumors in the GSE35988 dataset and with normal tissue in the GSE38241 dataset led to large separation between the PLS scores of the two classes (**Figures 4-6C, 4-6D**). The leave-one-out cross-validation accuracy was 100% for both of these comparisons.

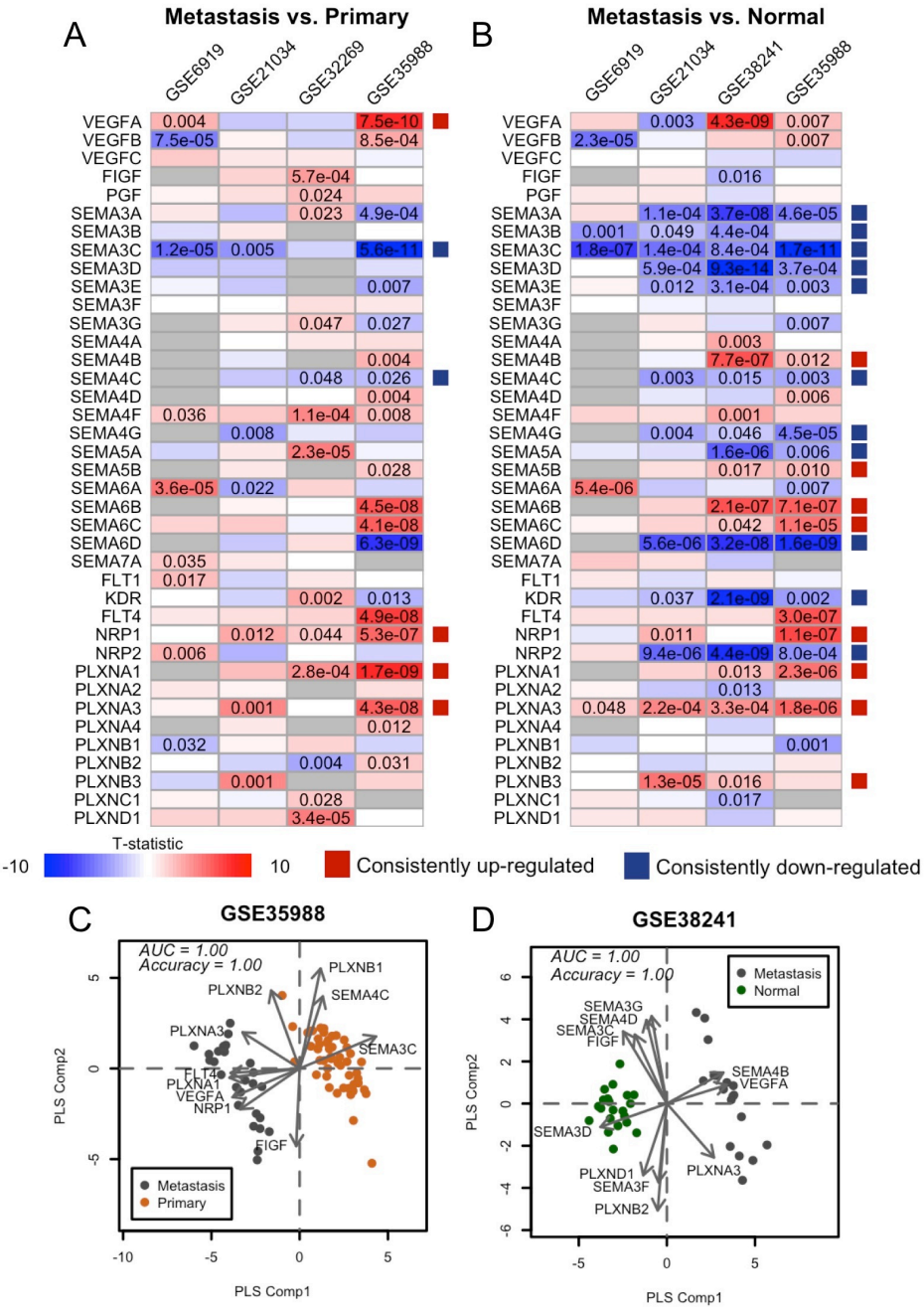

Figure 4-6: Pro-angiogenic VEGF/Sema gene expression in prostate cancer metastasis.

A-B: A greater number of VEGF/Sema genes have recurrent expression alterations when comparing metastases to normal samples (B) than to primary tumor samples (A). The p-values are displayed according to the same criteria as in **Fig.1**, and red and blue boxes on the right hand sides of panels A and B indicate recurrent up- and down-regulation, respectively. **C-D:** PLS-DA

scores plots show separation of metastases and primary tumors in GSE35988 (C) and of metastases and normal tissues in GSE38241 (D). Each dot represents a sample with colors as indicated. Arrows correspond to gene loadings in the PLS-DA models, with the names of the genes displayed in the vicinity of the arrowhead. Only the genes with the largest magnitude loadings vectors are displayed. Accuracy refers to the accuracy of the LOOCV predictions; AUC refers to the area under the curve of the LOOCV ROC curve. In both cases, values of 1 indicate perfect prediction.

4.3.4 VEGF/Sema alterations are consistent across multiple metastases within individual patients

The metastasis samples in the GSE38241 dataset included three to four metastases per patient from five different patients, providing the opportunity to analyze VEGF/Sema gene expression in both inter- and intra-patient contexts. We observed a high degree of consistency between metastases from the same patient, with *VEGFA* consistently up-regulated and several class 3 semaphorins, *KDR*, and *NRP2* consistently down-regulated (**Figure 4-7A**). Some genes that were not significantly altered when comparing all metastases to normal samples did show patient-specific alterations: *VEGFC* and *SEMA6A* were up-regulated in some metastases and down-regulated in others. *SEMA6A* was consistent within patients, while *VEGFC* was not.

To assess the overall consistency of metastases from individual patients, we performed consensus *K*-means clustering of the metastatic samples and compared the clusters to the patients of origin. Repeating the clustering on random subsets of the data yielded a consensus matrix with clear separations of the groups when the number of clusters was five (**Figures 4-7B, 4-7C**). This did not perfectly separate the patients, but there was a significant association between clusters and patients (χ^2 test P=0.001). We also found that if we expanded our set of VEGF/Sema genes to include other ligands and receptors important in angiogenesis, we obtained consensus *K*-means clusters that resulted in perfect separation of the patients into clusters (**Figures 4-7D, 4-7E**). This supports the hypothesis that, although signaling pathways that contribute to angiogenesis may vary from patient to patient, the multiple metastases that can arise from a single prostate tumor would likely all respond (or not respond) to the same or similar therapies.

4.3.5 Computational modeling of the VEGF/Sema pathway stratifies patients

To further analyze the contributions of gene expression changes to angiogenesis-related signaling, we developed a computational model consisting of ordinary differential equations

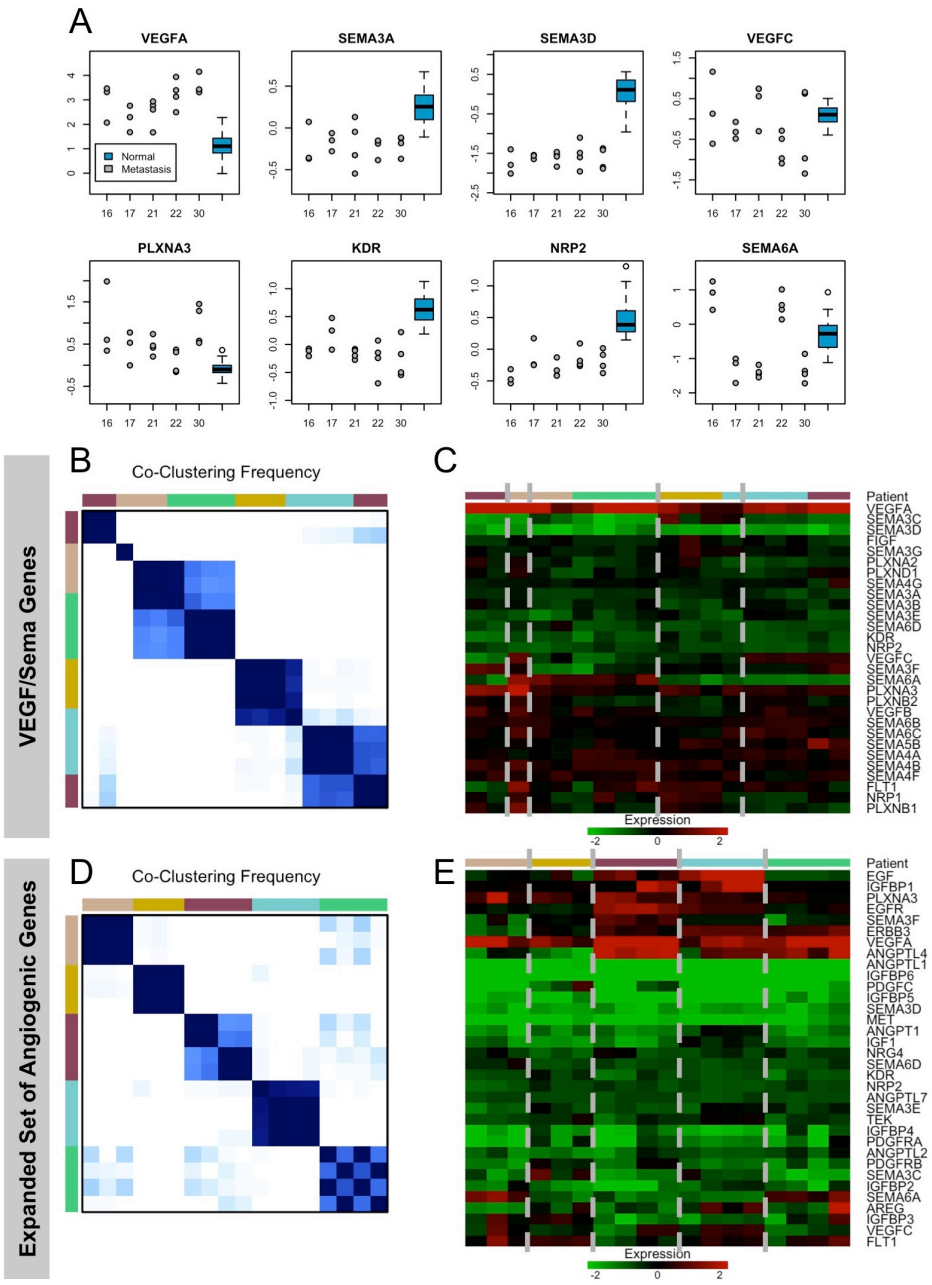

Figure 4-7: Intra-patient variability of metastases is low relative to inter-patient variability.

A: Many genes have consistent alterations in metastatic samples from the same patient while variable expression patterns between patients are more common. Blue box plots indicate the range of expression in the 21 normal samples, with the upper and lower ends of the boxes corresponding to the third and first quartiles of the data, respectively. The numbers along the *x*-axis indicate to which patient the dots above correspond. **B:** Consensus *K*-means clustering of

GSE38241 metastases according to VEGF/Sema expression shows a consistent co-clustering pattern for 5 clusters. Dark blue indicates a high frequency of co-clustering between consensus runs, while white indicates no co-clustering. Colors at the top and left indicate the patient from which each metastatic sample was taken. **C:** Heatmap of gene expression across 18 GSE38241 metastases. The mean expression of each gene in the normal prostate tissue samples is subtracted from the expression of each gene in the metastases so that the heatmap shows up/downregulation relative to normal. Only the most variable of the VEGF/Sema genes are displayed. Dashed lines separate clusters, and colors at the top correspond to patients as in B. **D-E:** As for B and C, but with the set of 85 angiogenesis-related genes.

(ODEs) describing the ligand-receptor binding kinetics of the five VEGF ligands and seven class 3 Semaphorins. A detailed model is essential, given the high inter-individual variability in gene expression, and given that predicting the outcome on signaling of simultaneous changes in expression of multiple genes encoding competing ligands and receptors becomes difficult. For a more detailed description of the model development, see Section 4.1.6. A key component of the model is that it incorporates both tumor cells and tumor endothelial cells, and these cells can express both the ligands and receptors of the VEGF and Sema pathways. We show some general characteristics of this model in **Figure 4-8**. On both cell types, VEGF RTKs and Plexin receptors tend to be present either in a ligated or neuropilin-coupled (or both) state (**Figure 4-8A, 4-8B**). Neuropilins, which outnumber the other receptors (**Table 4-6**), are present mostly in an uncoupled, unligated state. The effect of individually doubling the production rate of each protein in the model shows how each protein affects each receptor signaling complex (**Figure 4-8C, 4-8D**).

Instead of using an 'average' model, the tumor cell ligand and receptor production rates were varied based on the gene expression data in the GSE35988 dataset, which included normal prostate, primary prostate tumors, and metastatic tumors. This enabled us to run many simulations – one for each individual – and to predict the amounts of VEGF and Semaphorin signaling complexes in a patient-specific manner.

The results of simulating many individuals with tumors were divided into groups (benign, localized, metastatic) and the distribution of predicted signaling outputs across the population are calculated. Binding of the two major isoforms of VEGFA ($VEGF_{165}$ and $VEGF_{121}$) to VEGFR-2 and VEGFR-1 on endothelial cells was lowest in primary (localized) tumors, whereas the collective binding of the class 3 Semaphorins to endothelial plexins decreased in primary tumors relative to normal tissue and decreased further in metastatic tumors (**Figure 4-9A**). The receptor binding trends closely followed ligand expression levels, with *VEGFA* expression being the

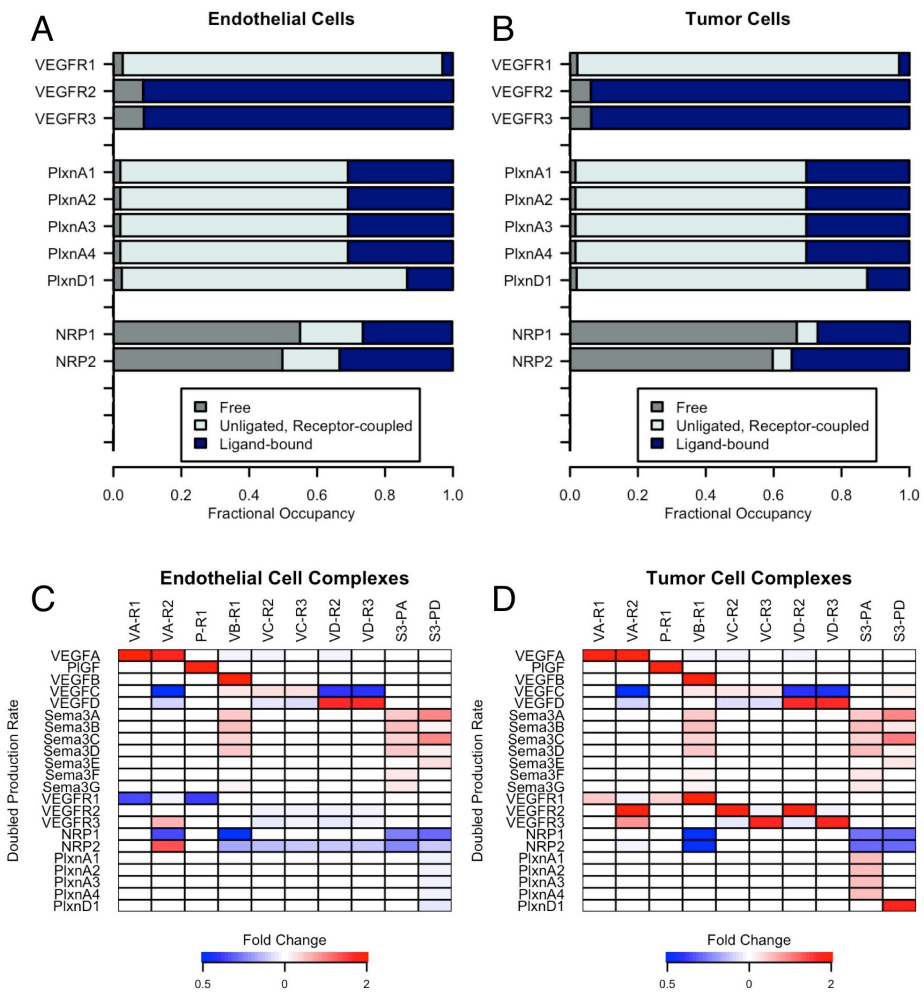

Figure 4-8: Receptor occupancy and sensitivity to protein production rates.

A-B, Most VEGFR and Plexin receptors are present in an unligated, Neuropilin-coupled state on both endothelial (A) and tumor (B) cells. Two exceptions are VEGFR-2 and VEGFR-3, which do not form ligand-independent complexes with Neuropilin. **C-D,** The production rate of each of the 22 proteins was doubled while holding other protein production rates steady. The heatmap shows the $\log_2$ ratio of the post-doubling steady-state to pre-doubling steady-state level of each receptor complex. VA-R1 includes all complexes with both VEGF-A and VEGFR-1 ($VEGFA_{165}$-R1, $VEGFA_{121}$-R1, and $VEGFA_{121}$-R1-NRP1), and so on for the rest of the complexes listed along the top of the figure. Receptor production rates were doubled on tumor cells only; thus, the effects on endothelial (C) and tumor (D) cells varied.

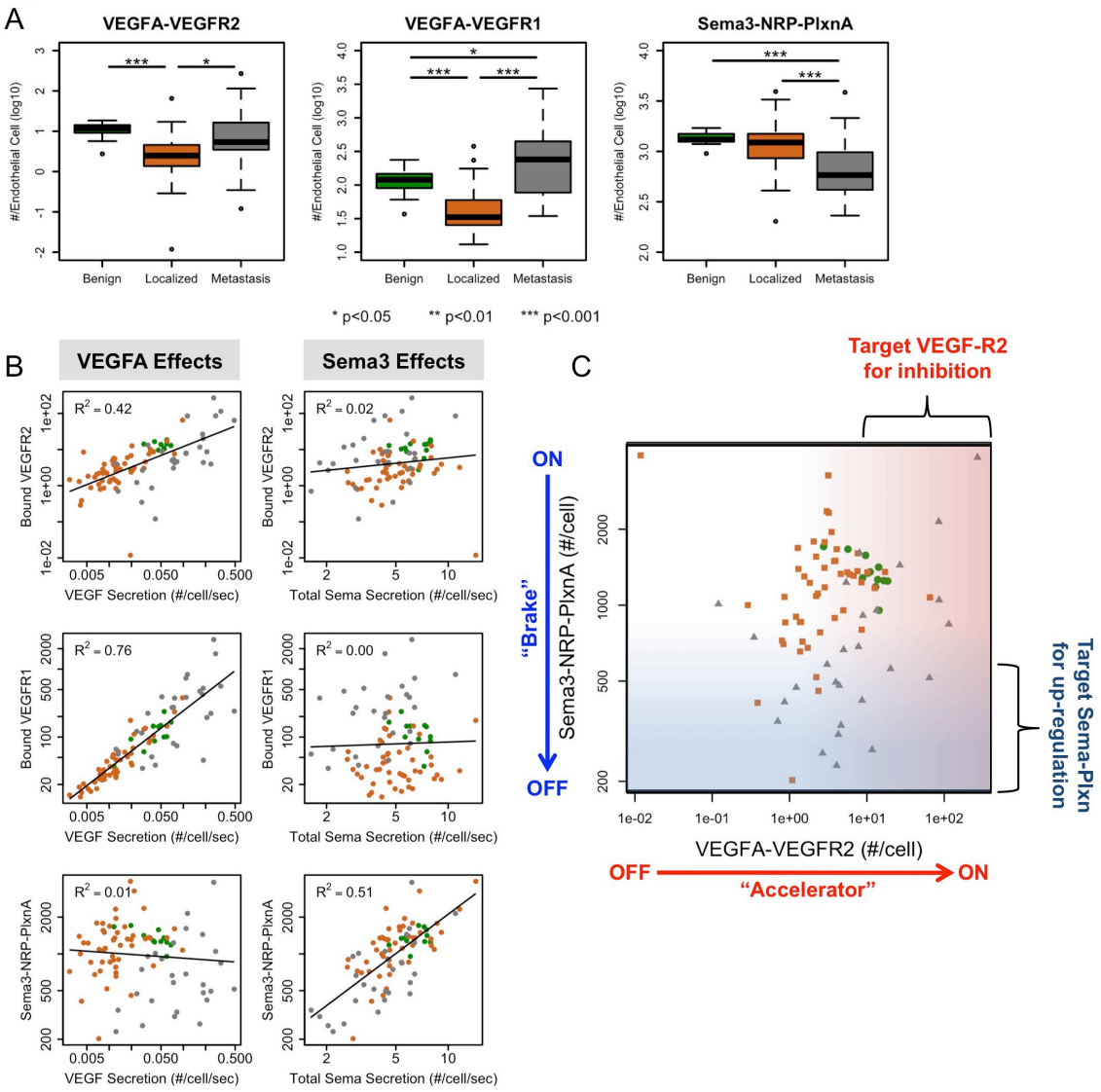

Figure 4-9: Simulated VEGF/Sema ligand-receptor binding on endothelial cells.

A: VEGFR2 (left), VEGFR1 (middle), and Sema3-NRP1 (right) binding in benign prostate (n=12), primary tumors (n=49) and metastatic tumors (n=27) in the GSE35988 dataset. **B:** VEGF secretion and total Sema3 secretion most strongly affect their respective ligand-receptor complexes although weak competitive effects are observed. Lines represent least squares fits of the log-transformed simulated receptor binding data to the gene expression data. R^2 values represent the proportion of variance explained by the least squares fit. **C:** Scatter plots of simulated VEGFA-VEGFR2 and Sema3-NRP-PlxnA across tissue types show that only fraction of the metastatic samples fall into the expected anti-VEGFA responsive region, i.e. high VEGFA-

VEGFR2 signaling and low Sema3-NRP-PlxnA signaling. Gradients correlate with expected favorability for angiogenesis: darker red for higher VEGFA-VEGFR2 and darker blue for lower Sema3-NRP-PlxnA. Colors indicate tissue type: Benign (green); Localized (orange); Metastasis (gray).

predominant factor driving VEGFR-1/2 binding, while the total Sema3 expression accounted for most of the variation in Sema3-NRP-PlxnA binding (**Figure 4-9B**). The receptor binding profiles on tumor cells (**Figure 4-10**) were similar to those on endothelial cells. These results indicated that primary tumors are associated with both pro-angiogenic (decreased Sema3-NRP-PlxnA binding) and anti-angiogenic (decreased VEGFR-2 binding) alterations, making it difficult to predict whether primary tumors would benefit from therapies that inhibit VEGF signaling. On the other hand, the alterations in metastatic tumors were all pro-angiogenic: VEGFR-2 binding was higher and Sema3-NRP-PlxnA binding was lower. The range of VEGFR-2 binding was highest in metastatic samples, going beyond both the low and high ends of the range of normal samples. This suggested an important role for patient selection in the use of anti-angiogenic therapies, as only the patients with high baseline VEGFR-2 signaling would be expected to respond.

Aside from anti-angiogenic signaling initiated by Sema3 binding to neuropilins and plexins, Sema3s also may inhibit angiogenesis by competitively displacing VEGF from neuropilins. In our model, we found only a weak competitive effect: the least squares fits for the non binding ligand-receptor pairs (VEGFA effects on Sema3-NRP-PlxnA and Sema3 effects on VEGFA-VEGFR-2 in **Figure 4-9B**) had slight negative slopes, but the least squares models explained a very small proportion of the overall variance. This was due to the fact that the amount of NRP1 and NRP2 present in ligand-containing signaling complexes was low relative to the total amount of neuropilins present on endothelial cells. If the amount of endothelial NRP1 and/or NRP2 were lower or the ligand secretion rates were higher, competition would be expected to have more impact. Thus, although direct VEGF-Sema competition for neuropilin is included in the model, the quantitative impact of the competition is predicted to be small.

Our simulation results suggest that determining patients with sensitivity to anti-angiogenic therapies would require biomarkers consisting of multiple predictor variables. Each individual tumor has a different predicted level of VEGFA-VEGFR-2 and low Sema3-NRP-

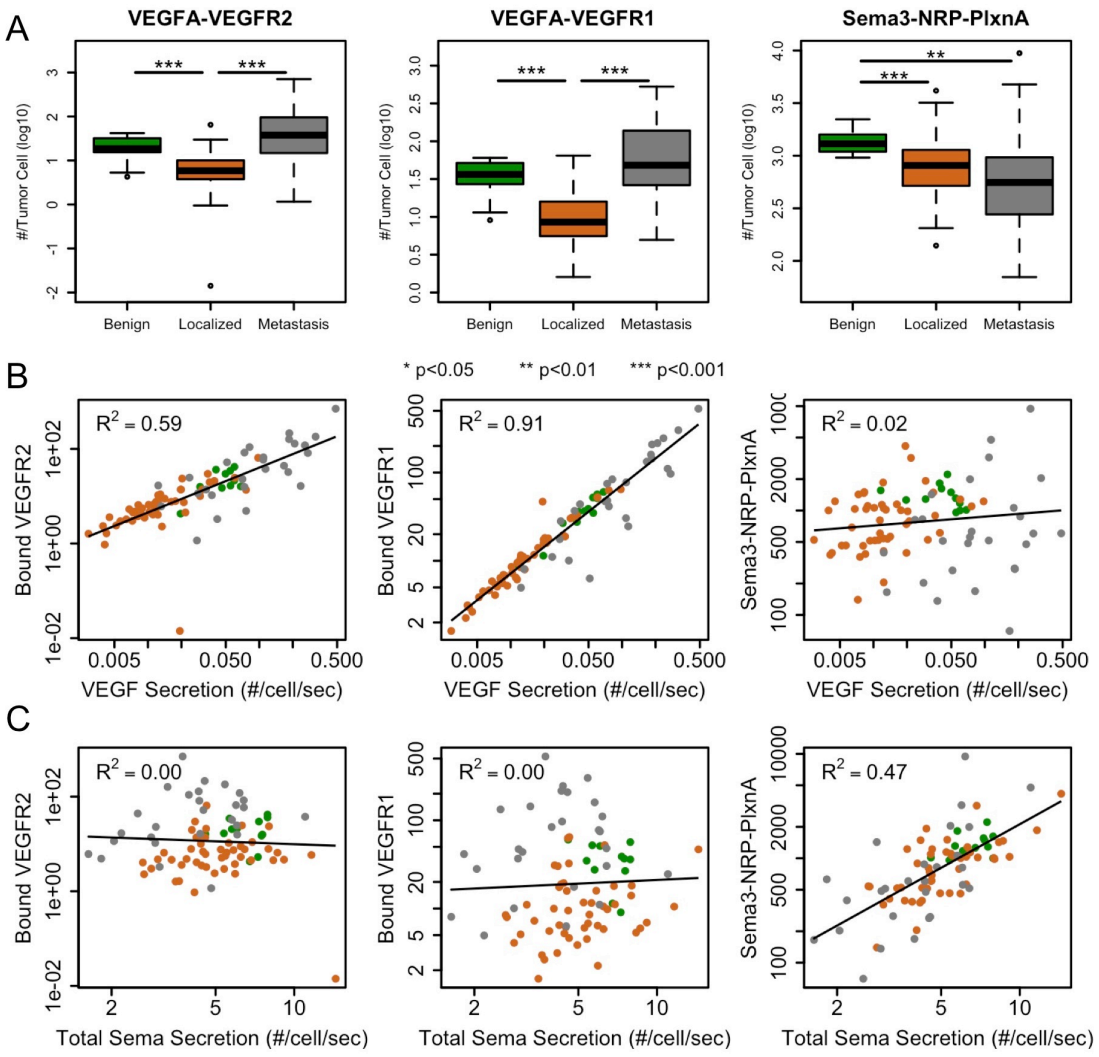

Figure 4-10: Simulated VEGF/Sema ligand-receptor binding on tumor cells.

A, Box plots of binding by tissue type (benign, localized, metastasis). **B-C,** Scatter plots showing the effects of VEGF secretion (B) and total Sema3 secretion (C) on binding.

PlxnA activity (**Figure 4-9C**). The red and blue gradients, and the arrows on the axes, indicate direction of increasing pro-angiogenic signaling for VEGFR2 (accelarator 'ON') and PlxnA1 (brake 'OFF'), respectively. Metastatic samples predominantly have high simulated VEGFA-VEGFR-2 and low Sema3-NRP-PlxnA (lower right quadrant of **Figure 4-9C**: accelerator 'ON', brake 'OFF'), and would be expected to benefit the most from VEGF-targeting therapies. However, this is not true of all metastatic tumors; a minority of patients with high VEGFA-VEGFR-2 also have high Sema3-NRP-PlxnA, possibly negating any clinical benefit of an anti-VEGFA agent. That combination is also predicted to be the most common for the benign tumors (green symbols in **Figure 4-9C**). There are also a subset of metastatic tumors predicted to have low VEGFA-VEGFR2 signaling. This heterogeneity amonst the metastatic tumors reinforces the need to bring an individualized understanding to therapeutic selection. For those with localized disease (grey symbols in **Figure 4-9C**) the characteristic signaling state is low VEGFA-VEGFR2 and high Sema-Plexin (upper left quadrant of **Figure 4-9C**: accelerator 'OFF', brake 'ON'), correlating with low angiogenesis potential and no metastases. Further analysis of this model and availability of both gene expression and clinical anti-VEGF outcome data will allow us to develop simulation-based biomarkers.

4.4 Discussion

As with many types of cancer, angiogenesis may enable prostate cancer growth and progression. Tumor vessels tend to be more irregular than normal prostate vessels [139,140], and higher microvessel density is associated with higher tumor stage [141,142]. Some studies have reported a reduction in tumor microvessel density relative to normal prostate tissue, but even in these studies it is noted that “hotspots” exist i.e. regions with locally high microvessel densities [140]. Prostate tumors are known to have intra-tumoral heterogeneity [121,143], raising the possibility that these hotspots correspond to regions with a pro-angiogenic genomic signature. Multiple molecular regulators of angiogenesis could be responsible for the tumor-associated

changes in the microvasculature. The predominant pro-angiogenic factor, VEGFA, is elevated in the plasma of patients with metastatic prostate cancer, but not in patients with localized primary prostate tumors [144,145,146,147,148]. This does not rule out a role for VEGFA in primary tumors, as immunohistochemistry has revealed increased VEGFA in primary tumors relative to benign prostate tissue, and in castration-resistant tumors relative to hormone-naïve [149]. Our analysis of previously published gene expression data in this study revealed increases in *VEGFA* expression only in metastatic tumor samples; *VEGFA* expression in primary tumors was actually decreased. Despite elevations in VEGFA in metastatic disease, most VEGF inhibitors have failed clinical trials in metastatic CRPC [31,150,151,152]. One compound that has shown some promise in phase II trials, cabozantinib, also targets the Met receptor [153]. The lack of success of VEGF inhibitors suggests that other angiogenesis modulators may be involved. Therefore, we performed an analysis of VEGF-related genes and a family of potential modulators, the semaphorins, across stages of prostate cancer to gain a system-wide perspective on VEGF activity in this disease. Our methodology could be widened in the future to include more relevant RTK families.

An active area of research in prostate cancer molecular biomarkers is predicting whether a primary tumor will eventually become aggressive or if it will remain indolent as indolent prostate cancers can often be left untreated and monitored. We found that a multivariate VEGF/Sema signature was associated with aggressive tumors. We used PLS-DA for distinguishing between aggressive and indolent tumors, as well as different tissue types, as it provided both effective classification and information about the correlation structure within the gene expression data. This allowed for interpretation of the expected joint effects of expression variation on VEGF signaling activity. We expected the performance of this approach to be comparable to other multivariate methods used for similar purposes, including linear discriminant analysis [154], random forests [155], and decision trees [156]. Our VEGF/Sema PLS-DA biomarkers had similar prognostic capabilities to PLS-DA biomarkers based on the genes from

other prostate cancer prognostic indicators available and in development for clinical use. We used PLS-DA, as opposed to the algorithm actually used in the prognostic indicators, because the algorithms and parameters used in these indicators were proprietary. The VEGF/Sema signature of aggressiveness that we found contained several genes that promote lymphangiogenesis, including *FIGF* (VEGFD), *NRP2*, *VEGFC*, and *KDR*. This fits with previous research that suggests lymphangiogenesis plays a role in allowing primary tumors to escape their tissue of origin via the lymphatic vessels [138,157,158]. Inhibiting one or more of these genes could be an effective therapeutic mechanism in patients whose tumors are predicted to be aggressive, possibly as a neoadjuvant therapy prior to radical prostatectomy.

Most approaches to targeting VEGF-dependent cancer angiogenesis have relied on inhibiting VEGF or VEGFR2. This focus on blocking the accelerator of tumor angiogenesis may have overlooked a key aspect of tumor angiogenesis – the endogenous brakes provided by the Semaphorins. Analysis of metastatic tumor gene expression suggested several possible reasons for the failure of anti-angiogenic therapies in prostate cancer. As noted above, *VEGFA* expression was consistently elevated in metastases relative to normal samples, but treatment of metastatic prostate cancer with VEGF inhibitors typically fails. Reductions in class 3 semaphorins were observed, which could enhance VEGF signaling by making more NRP1 available; lower Sema3 levels could also enhance angiogenesis by removing inhibitory signals mediated by plexins (i.e. the brakes on tumor angiogenesis are removed). Our simulation data provide mechanistic insight that supports this latter case: Sema3s did not appear to alter the availability of NRP1 to VEGF, but the formation of anti-angiogenic Sema3-Plexin complexes was decreased in metastases. Thus one possible mechanism of resistance to VEGF inhibitors is reduced anti-angiogenic Sema3 signaling. With the repression of anti-angiogenic Sema signaling, blocking the accelerator may be insufficient to halt the runaway tumor vasculature; thus treatment might best be achieved, or augmented, by restoring or replacing the endogenous brakes on tumor angiogenesis.

We observed high correlation among the various metastases from each individual (**Figure 4-7**). These multiple metastases arising from a single primary tumor may then respond similarly to particular therapies. Also, given recent findings of heterogeneity in primary tumors, it may suggest that either (a) metastases come from a selected subset of the primary tumor or (b) environments receiving and nurturing metastases may cause the metastatic tumor cells to converge. The gene expression clustering analysis in this study also presented other possible resistance mechanisms that we did not simulate. For example, *SEMA6A* was up-regulated in metastases from two patients and down-regulated in three. It has roles in angiogenesis, potentially though an interaction with VEGFR2 [115].

To move beyond the 'average patient' that is typically simulated by molecularly-detailed mechanistic computational models, we have integrated mechanistic simulations with high-throughput data to create a population of tumor models that can simulate variability in receptor activation and response to treatment. Our simulation results suggest that determining patients with sensitivity to anti-angiogenic therapies would require biomarkers consisting of multiple predictor variables. This is one of the most important considerations for this type of approach. Linear approaches such as PLS-DA are ultimately limited to identifying linear combinations of effects within the gene expression data. By adding the mechanism-based, quantitative, nonlinear protein-interaction network, we can generate latent variables that integrate both gene expression and mechanistic information. These predicted mechanistic latent variables may then be more predictive of the angiogenesis potential of the tumor and of the outcome of therapeutic inhibition.

Simulations of the kind presented here can be expanded to incorporate additional proteins and multiple tissue compartments to model drug pharmacokinetics and pharmacodynamics. This will be useful for developing patient-specific models capable of identifying the most appropriate molecular therapies. Therapies could include VEGF-targeting agents such as bevacizumab as well as drugs that target neuropilin, either by blocking ligand binding to neuropilin or by blocking the

coupling of neuropilin to other receptors. Inclusion of semaphorins in these models allows us to analyze whether neuropilin-targeting therapies could inadvertently have a pro-angiogenic effect due to reduced semaphorin signaling. Additional semaphorins (classes 4, 5, 6, and 7) could be added as data for the kinetics becomes available. Further development of the model could also address several of the limitations of the current model. Proteolytic processing is known to alter the receptor binding and therefore the activity of VEGFC, VEGFD, and several of the class 3 Semaphorins. We neglected these effects here due to a lack of data describing the relative amounts of processed and unprocessed forms in tumors. Additionally, we have assumed that competition of VEGF and class 3 Semaphorins occurs due to the inability of both to occupy neuropilin receptors simultaneously. Further validation against published *in vitro* data may allow us to refine this assumption for specific VEGF and Semaphorin ligands; there is evidence that certain ligands (and proteolytically processed forms) are more or less able to sterically inhibit the binding of other ligands to neuropilin.

Several limitations of gene expression data may affect the conclusions of this study. Prostate tumors are typically multifocal and heterogeneous [159,160], creating the possibility of gene expression data for a patient that does not reflect the clone that actually gives rise to aggressive disease. This is a possible explanation for the difference between the TCGA and GSE21034 datasets in the VEGF/Sema signatures found in this study to distinguish aggressive and indolent tumors. Additionally, gene expression levels are not always representative of the level of the corresponding protein. Instead of equating gene expression to protein levels, we assumed that protein secretion *rates* were proportional to gene expression in our patient-specific simulations. This is a useful first approximation; further analysis could include miRNA data and other regulatory factors that would be expected to influence the rate of translation of a transcript to a protein. A final limitation is the small size of metastasis datasets. Metastatic samples are of particular interest because metastases give prostate cancer its lethality and here we show that they

may have increased VEGF signaling activity. A larger rapid autopsy dataset would allow expansion of limited analysis performed here.

5 Cancer-Specific Differences in Sensitivity to VEGF-Targeting Therapies: Predictions from Expression-Based Personalized Models

5.1 Summary

There are currently no predictive biomarkers of VEGF-targeting therapies that can accurately distinguish responders from non-responders. A major obstacle in developing biomarkers is the difficulty in obtaining measurements of the levels of VEGF signaling complexes in tumor patients. Here we develop a computational modeling framework that incorporates patient gene expression data in a multiscale compartment model of VEGFA and PlGF signaling. The model allows us to simulate the effects of two VEGF-targeting therapies in virtual populations of breast, kidney, and prostate cancer patients. We used this simulated dataset to define treatment response metrics consisting of aggregates of highly correlated variables that were altered substantially by either treatment. The typical pattern observed was a reduction in VEGF-receptor binding in the tumor accompanied by elevated VEGF concentrations in the plasma and normal tissues. Varying effects on PlGF binding were observed: PlGF1 binding was elevated in the tumor in response to anti-VEGF, whereas binding of both isoforms was elevated in tumors in response to anti-NRP1. After defining metrics, we analyzed the performance of various biomarkers as predictors of these response metrics. The possible biomarkers included both gene expression data and simulated baseline data. We found that simulated baseline data almost always served as better predictors of treatment response metrics. This suggested that the levels of various protein complexes, which may be difficult to measure *in vivo*, but which may be simulated based on other measurable parameters, may be better predictors of treatment response.

5.2 Methods

5.2.1 Compartment modeling

To examine the effects of gene expression variability and drug interactions, we simulated the binding and transport of ligands throughout three compartments of the body: the tumor, the blood, and all other normal tissues. Compartment models of this kind have previously been used to analyze the distribution of VEGFA isoforms in diseased tissues [44,47]; here, we expand previous compartment models to include two isoforms of PlGF. The parameters for all rate constants in the reactions between ligands and receptors were as listed in **Tables 4-3** and **4-4** (note that here, only VEGFA and PlGF were present).

The ordinary differential equations describing association and dissociation of ligands and receptors, as well as coupling between receptor complexes, remain the same as in Section 4. Additional terms are included to describe the transport of soluble proteins and protein complexes between compartments. Three types of transport reactions are included in the model: vascular permeability, lymphatic transport, and clearance from the blood. **Equation 5-1** is an example of the how these new terms affect the differential equation for $VEGFA_{165}$ in the blood.

Equation 5-1: Intercompartment transport of $VEGF_{165}$

$$\frac{d[V_{165}^B]}{dt} = -c_V[V_{165}^B] - k_{p,V}^{NB}\frac{S_{NB}U_N}{U_B}\left(\frac{[V_{165}^B]}{K_{AV}^B} - \frac{[V_{165}^N]}{K_{AV}^N}\right) - k_{p,V}^{TB}\frac{S_{TB}U_T}{U_B}\left(\frac{[V_{165}^B]}{K_{AV}^B} - \frac{[V_{165}^T]}{K_{AV}^T}\right) + \frac{k_{L,V}^N}{U_B}\frac{[V_{165}^N]}{K_{AV}^N}$$

The *B*, *N*, and *T* superscripts in this equation indicate the compartment (blood, normal, or tumor). The vascular permeability constants $k_{p,V}$ and the surface area-to-volume ratios *S* have two compartment superscripts or subscripts, indicating that transport occurs across the vascular endothelium separating the blood from either the tumor or the normal tissue. The values for these constants and the other constants (clearance rate constant c_V, compartment volumes *U*, available interstitial volume fractions K_{AV}, and lymphatic transport rate constant k_L) are given in **Table 5-1**.

Table 5-1: Transport and geometric parameters.

Parameter	Symbol	Value	Units
Microvessel SAV (tumor compartment)	S_{TB}	105	cm^2/cm^3 tissue
Tumor SAV	S_T	1416	cm^2/cm^3 tissue
Microvessel SAV (normal compartment)	S_{NB}	108	cm^2/cm^3 tissue
Parenchyma SAV (normal compartment)	S_P	664	cm^2/cm^3 tissue
Interstitial volume fraction (tumor)	K_{AV}^T	0.611	cm^3/cm^3 tissue
Interstitial volume fraction (blood)	K_{AV}^B	0.567	cm^3/cm^3 tissue
Interstitial volume fraction (normal)	K_{AV}^N	0.0816	cm^3/cm^3 tissue
Tumor compartment volume	U_V	940	cm^3 tissue
Blood compartment volume	U_B	5,000	cm^3 tissue
Normal compartment volume	U_N	61,200	cm^3 tissue
Endothelial cell surface area	a_{EC}	$1 \cdot 10^{-5}$	cm^2/cell
Parenchyma/Tumor cell surface area	a_{TC}	$1.85 \cdot 10^{-5}$	cm^2/cell
VEGF normal tissue permeability	$k_{p,V}^{NB}$	$4 \cdot 10^{-8}$	cm/sec
VEGF tumor tissue permeability	$k_{p,V}^{TB}$	$4 \cdot 10^{-7}$	cm/sec
sVEGFR1 normal tissue permeability	$k_{p,R1}^{NB}$	$1.9 \cdot 10^{-8}$	cm/sec
sVEGFR1 tumor tissue permeability	$k_{p,R1}^{TB}$	$1.9 \cdot 10^{-7}$	cm/sec
Drug normal tissue permeability	$k_{p,D}^{NB}$	$3 \cdot 10^{-8}$	cm/sec
Drug tumor tissue permeability	$k_{p,D}^{TB}$	$3 \cdot 10^{-7}$	cm/sec
Clearance – VEGF and PlGF	c_V	$1 \cdot 10^{-3}$	sec^{-1}
Clearance – sR1 & sR1 complexes	c_{R1}	$5 \cdot 10^{-6}$	sec^{-1}
Clearance – drug & drug complexes	c_D	$3.2 \cdot 10^{-7}$	sec^{-1}
Lymphatic drainage – normal tissue	k_L^N	0.033	cm^3/sec

5.2.2 Angiogenesis-targeting therapies

We tested the effects of two therapies in the model: an antibody that binds to VEGFA and an antibody that binds to the co-receptor Neuropilin-1. The anti-VEGF antibody binds to $VEGFA_{165}$ and $VEGFA_{121}$ and prevents them from binding to receptors. The anti-NRP1 antibody binds to NRP1 and prevents it from coupling with other receptors but allows ligand binding. Both therapies were infused into the blood over an hour and a half to final doses of 10 mg/kg (all patients were equal mass – 70 kg). A duration of three weeks was then simulated to observe the effects of the treatments. Equations 5-2 and 5-3 are examples of how these drugs effect the concentrations of $VEGFA_{165}$ in the blood and free NRP1 on the tumor endothelium.

Equation 5-2: Bevacizumab interactions in the blood.

$$\frac{d[Beva^B]}{dt} = -k_{on}^{V-Beva}[V_{165}^B][Beva^B] + k_{off}^{V-Beva}[V_{165} - Beva]^B$$
$$-k_{on}^{V-Beva}[V_{121}^B][Beva^B] + k_{off}^{V-Beva}[V_{121} - Beva]^B$$

Equation 5-3: Anti-NRP1 interactions on the tumor endothelium.

$$\frac{d[N1Ab^T]}{dt} = -k_{on}^{N1-N1Ab}[N1Ab^T][NRP1_{Endo}^T] + k_{off}^{N1-N1Ab}[N1Ab - NRP1]_{Endo}^T$$
$$-k_{on}^{N1-N1Ab}[N1Ab^T][V_{165} - NRP1_{Endo}^T] + k_{off}^{N1-N1Ab}[N1Ab - V_{165} - NRP1]_{Endo}^T$$
$$k_{on}^{N1-N1Ab}[N1Ab^T][PlGF2 - NRP1_{Endo}^T] + k_{off}^{N1-N1Ab}[N1Ab - PlGF2 - NRP1]_{Endo}^T$$

5.2.3 Virtual population

To create a virtual population of models, we used gene expression data to vary the production rate of proteins by tumor cells. We made the assumption that all variability in gene expression was due to tumor cells; thus, the production rates in the model varied in the tumor cells while the production rate of receptors on endothelial cells remained constant across the population.

We varied protein production rates linearly with untransformed gene expression data. We set the nominal tumor protein production rates for VEGFA, PlGF, VEGFR-1, VEGFR-2, and NRP1 by finding the production rates that corresponded to 2.67 pM total VEGFA, 0.71 pM total PlGF, and 120 pM sVEGFR1 in the blood. These values correspond to the median concentrations found in cancer patients The production rates that yielded these concentrations were therefore the population median production rates. To determine the patient-specific production rates, we normalized gene expression data by dividing by the median. We then multiplied this quantity by the corresponding median protein production rate to get the individual production rate for each gene/protein. Although this is by no means an exact method for converting between gene

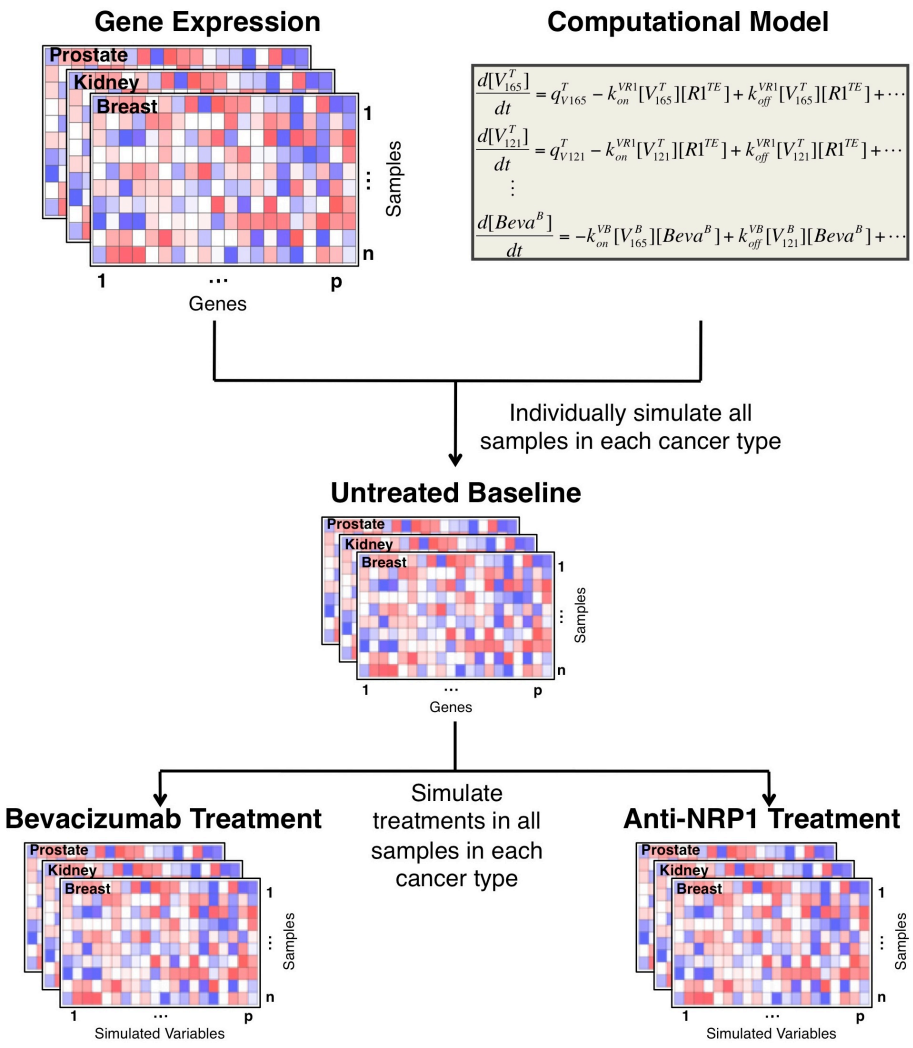

Figure 5-1: Flowchart of VEGF compartment model simulations for TCGA data.

Gene expression data from three cancer types were used to parameterize production rates of $VEGFA_{165}$, $VEGFA_{121}$, PlGF1, PlGF2, sVEGFR1, VEGFR1, VEGFR2, and NRP1 in tumor cells. These individualized models were first simulated in the absence of drugs to find the untreated baseline. Then, either anti-VEGF (bevacizumab) or anti-NRP1 were added via infusion into the bloodstream. This resulted in three simulated datasets: the untreated baseline, anti-VEGF-treated, and anti-NRP1-treated. Subsequently, biomarkers for predicting the behavior of the two treatment datasets were developed based on either the gene expression or the untreated baseline data.

expression and protein production, it is an approximation that preserves a linear relationship between protein production and gene expression, which does occur in some cases.

We used TCGA RNA-Seq datasets to introduce variability to protein production rates. We focused on breast adenocarcinoma, renal cell carcinoma, and prostate adenocarcinoma. For each cancer type, we generated three simulated datasets: one with no treatments, one with anti-VEGF (bevacizumab), and one with anti-NRP1 (**Figure 5-1**).

5.3 Results

5.3.1 *VEGFA* and *PGF* expression drive VEGF- and PlGF-containing complexes.

The virtual population, with patient-specific protein production rates, was simulated in the absence of treatments until a steady state was reached. This corresponded to the untreated baseline present in tumors at diagnosis. For each cancer type, we applied principal component analysis to determine patterns of co-variability among the steady-state variables consisting of the concentrations of the ligands and receptor complexes throughout the three tissues in the model. We found that the principal component loadings associated with the first and second components always consisted entirely of VEGF- and PlGF-containing complexes, respectively (**Figures 5-2A, 5-2B,** and **5-2C**). Some variations in this pattern were observed: the number of VEGF-containing complexes with large magnitude first component loadings varied between cancer types, with renal cell carcinoma having more complexes with large loadings than breast or prostate cancers.

The pattern observed in the PCA loadings plots indicated that the variability in the steady-state simulated data was primarily driven by the expression of genes encoding the ligands, *VEGFA* and *PGF*, but not the receptors. Indeed, the Pearson correlations between the log-transformed gene expression and log-transformed level of signaling complexes on the tumor endothelium were greater than 0.9 for both *VEGFA* and *PGF* across the tumor types (**Figures 5-2D** through **5-2I**).

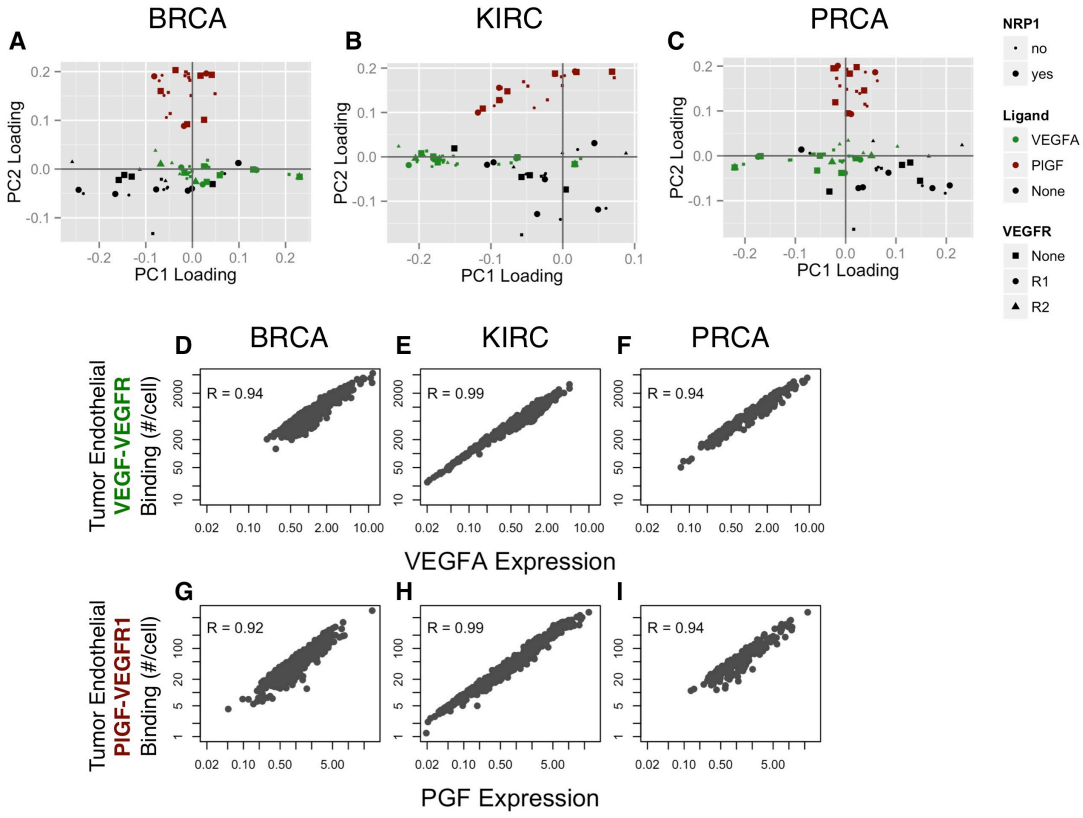

Figure 5-2: Steady-state behavior is driven by expression of the two ligands.

A-C, PCA of the steady-state simulation data for breast (A), kidney (B), and prostate (C) cancers reveals a pattern of correlation among VEGFA-containing as well as PlGF-containing complexes. The loadings plots show the projections of the 324 simulated variables onto the first two principal components. Red points correspond to any ligand or receptor complex that contains either PlGF isoform, while the green points correspond to any ligand or receptor complex that contains either VEGFA isoform. The size of the point indicates whether the variable is a NRP1-containing complex (large) or not (small). The shape of the point indicates if the variable is a complex containing VEGFR1 (circle), VEGFR2 (triangle), or neither (square). **D-F,** VEGFA expression is correlated with the aggregate steady-state level of all VEGFA-VEGFR complexes in breast (D), kidney (E), and prostate (F) cancers. **G-I,** Same as D-F, but with PlGF instead of VEGFA.

5.3.2 Derivation of treatment response metrics from PCA.

We simulated the intravenous infusion of two monoclonal antibody drugs in the virtual patient populations: anti-VEGF (bevacizumab) and anti-NRP1. To show the variability across each cancer population, we used the concentration of free VEGFA (both isoforms – $VEGFA_{165}$ and $VEGFA_{121}$) in the plasma, which is known to increase in many patients after administration of anti-VEGF. In all three cancer types, plasma free VEGFA increased on average in response to both anti-VEGF and anti-NRP1 (**Figure 5-3A** through **5-3C** and **5-3G** through **5-3I**). There was an initial repression of plasma VEGFA after addition of anti-VEGFA, but this was reversed quickly in most of the patients (**Figure 5-3D** through **5-3F**). The range of the plasma VEGFA elevation was higher in response to anti-VEGFA than anti-NRP1. The effect of receptor-mediated clearance of the anti-NRP1 antibody can be seen by the reduction in plasma VEGFA at later time points (**Figure 5-3J** through **5-3L**).

A difficulty with analyzing the post-treatment simulated data was the large size of the dataset: there were 324 variables comprising the soluble and surface-bound molecular complexes across cell types and tissues. Each of these variables in turn had simulation data available at 523 time points spanning the three-week treatment. To facilitate the analysis of this large dataset, we used PCA for dimension reduction. This allowed us to combine highly correlated variables together so that we could consider fewer variables while still retaining all relevant variability in the dataset.

An example of the PCA-based dimensional reduction follows for the response of the virtual populations to bevacizumab (**Figure 5-4A** through **5-4C**). We evaluated the time integral of the level of each complex over the three-week treatment duration and then computed the relative change in this quantity from the baseline steady-state level, which corresponds to time zero. This resulted in a simulated dataset comprising 324 variables for each patient in the three virtual populations, which was used as input for PCA. The loadings of the first four components were used to find sets of correlated variables. Soluble and receptor-bound VEGFA in the normal

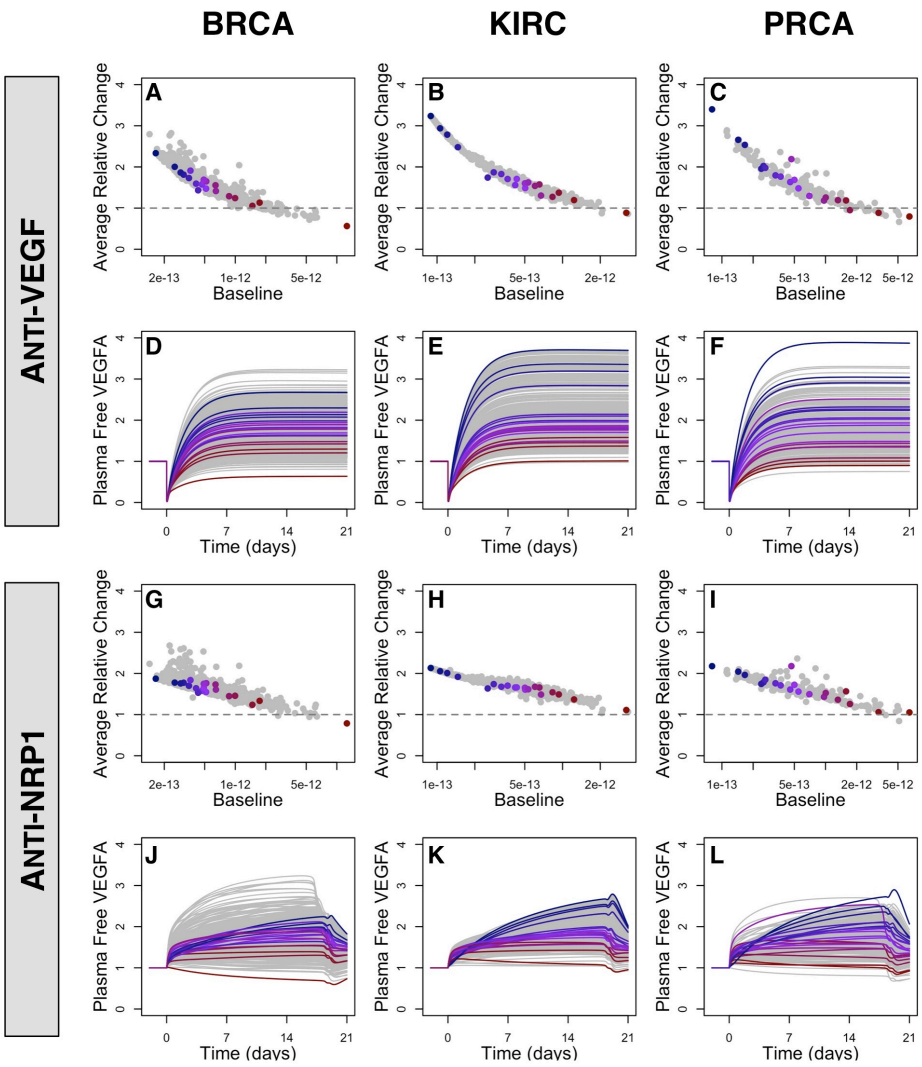

Figure 5-3: Plasma free VEGFA varies widely in response to VEGF-targeting therapies.

A-C, The aggregated steady-state level of plasma $VEGFA_{165}$ and $VEGFA_{121}$ is inversely associated with the fold change in plasma $VEGFA_{165}$ and $VEGFA_{121}$ in response to anti-VEGF in breast (A), kidney (B), and prostate (C) cancers. Colored dots correspond to 20 patients, evenly spaced when ranking them from highest to lowest baseline plasma VEGFA. Colors progress from blue to purple to red, with red corresponding to the highest level of plasma baseline VEGFA. **D-F,** Time courses of plasma VEGFA after administration of anti-VEGF in all patients in the breast (D), kidney (E), and prostate (F) virtual populations. Line colors are patient-specific as in A-C. **G-L,** As in A-F, but the drug is anti-NRP1.

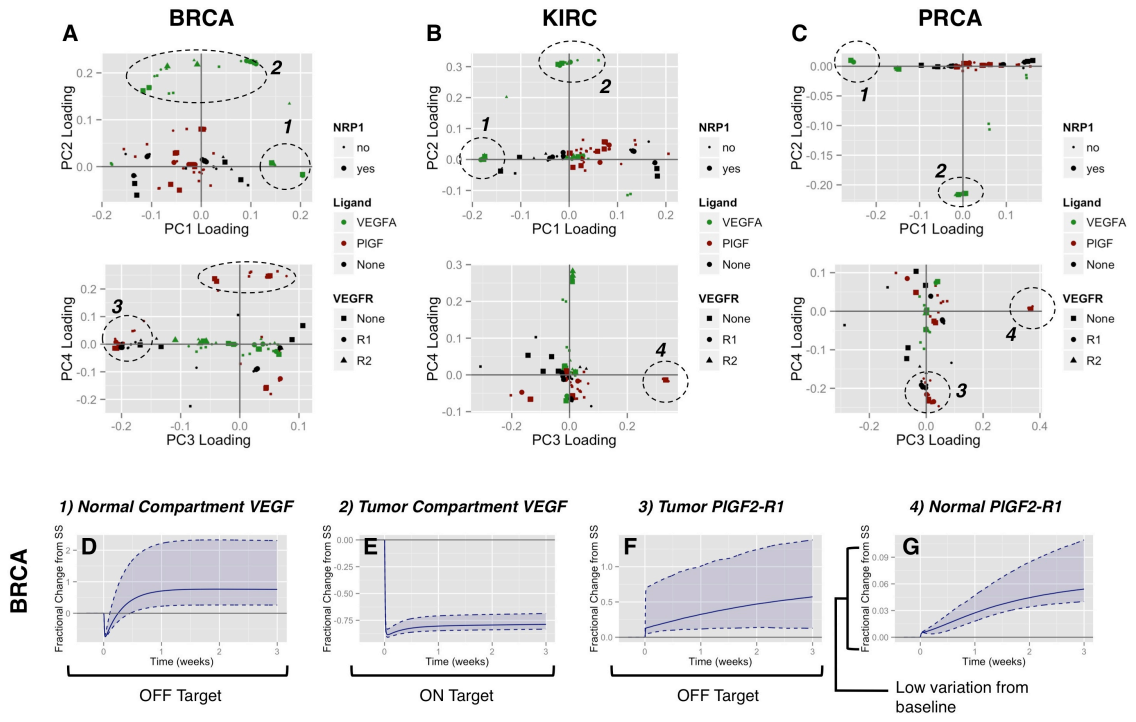

Figure 5-4: Derivation of bevacizumab response metrics.

A-C, PCA loadings plots show the relationships between simulated variables in response to bevacizumab for breast (A), kidney (B), and prostate (C) cancers. Patterns that recur in multiple tumor types are enclosed in dashed circles. The numbers correspond to the time course plots in D-G. **D-G,** Population time courses of aggregate variables consisting of the correlated simulated variables from A-C. Solid lines indicate the population median at each time point, while dashed lines correspond to the 5th and 95th percentiles at each time point. All curves are normalized by subtracting the sample-specific baseline value and then dividing by the baseline. Only breast cancer data is shown here; the kidney and prostate cancer data had very similar behavior. The drug response metrics found here correspond to on-target effects – reduced tumor compartment VEGF binding (E) – as well as off-target effects – elevated normal compartment VEGF binding (D) and tumor PlGF2-VEGFR1 binding (F). The elevation in normal tissue PlGF2-VEGFR1 binding (G) has only slight variation from the baseline level (<10%) thus it was not included as a metric in subsequent analyses.

compartment were associated with the first component in breast cancer (enclosed in a dashed circle and labeled *1* in **Figure 5-4A**), while the second component was associated with soluble and receptor-bound VEGFA in the tumor compartment (enclosed in a dashed circle and labeled *2* in **Figure 5-4A**). Isoform-specific effects were not noted; this was expected as bevacizumab binds both $VEGFA_{165}$ and $VEGFA_{121}$ with equal affinity. The third and fourth components for breast cancer were associated with PlGF-containing variables in an isoform-specific manner. Complexes containing PlGF in the tumor compartment were associated with the third component (enclosed in a dashed circle and labeled *3* in **Figure 5-4A**), while soluble and NRP1-bound PlGF2 in all compartments was associated with the fourth component (enclosed in a dashed circle but not labeled in **Figure 5-4A**).

Similar patterns were observed in kidney cancer (**Figure 5-4B**) and prostate cancer (**Figure 5-4C**). Normal compartment VEGF binding and tumor compartment VEGF binding were associated with the first and second principal components, respectively, in all three cancer types. A new set of variables was observed with a distinct loading pattern in kidney and prostate cancers: normal compartment PlGF2-VEGFR1 complexes (enclosed in a dashed circle and labeled *4* in **Figure 5-4B** and **5-4C**).

To aggregate the collections of variables with labels in **Figures 5-4A** through **5-4C**, we added together the simulated level of each variable at each time point, then computed the relative change from time zero. This method allowed us to preserve the relative cell surface densities of receptor complexes. The time courses of these aggregated relative variables showed the impact of bevacizumab on the breast cancer virtual population (**Figure 5-4D** through **5-4G**): VEGFA was redistributed from the tumor compartment to the normal compartment, and PlGF binding was enhanced, particularly for PlGF2 binding to VEGFR1 in the tumor and normal compartments. The range of variability across the population (dashed lines in time course plots in **Figure 5-4** indicate 5^{th} and 95^{th} percentiles, while solid lines indicate median) demonstrated a fairly consistent on-target response, namely reduction in tumor compartment VEGF binding (**Figure 5-**

4D). The increased level of normal compartment VEGF binding (**Figure 5-4E**) varied widely across the population, with some patients having negligible increases, while others had more than 200% increases. The elevations in PlGF2-VEGR1 binding in the tumor compartment (**Figure 5-4F**) also varied widely. We excluded PlGF2-VEGFR1 binding in the normal compartment (**Figure 5-4G**) from the bevacizumab metrics because it increased by less than 10%. The ranges of variation for the four aggregate variables in the other two cancer types (data not shown) was similar to the range in breast cancer.

Applying dimensional reduction to the anti-NRP1 responses yielded patterns with some similarities to the anti-VEGF responses. Normal and tumor compartment VEGF complexes each were strongly associated with principal components (data not shown). However, other patterns were difficult to discern, therefore we considered the metrics shown along the left side of **Figure 5-5**. These metrics are made up of some complexes including NRP1 and some that do not include NRP1. In breast cancer, addition of the anti-NRP1 antibody always increased the level of the complexes without NRP1 (first column of **Figure 5-5**) and always reduced the level of complexes with NRP1 (middle column of **Figure 5-5**). When we considered the net effect, we found that tumor VEGF binding decreased overall (**Figure 5-5F**) while the other three metrics increased (**Figure 5-5C, 5-5I, 5-5L**). All of these metrics had ranges of variability that exceeded 10%. The range of the changes was similar in the kidney and prostate populations as well (data not shown).

5.3.3 Prediction of drug sensitivity metrics.

Since clinical data describing patient responses to bevacizumab and anti-NRP1 was unavailable in these datasets, we used the drug metrics derived in the previous section. Our goal was to fit statistical models to these metrics using gene expression data or simulated baseline data as predictor variables. We tested all possible univariate linear regression models. We then found the best gene expression predictor for each metric and the best baseline simulation predictor for each metric (**Figure 5-6**). In most cases, the best simulated baseline variable had higher

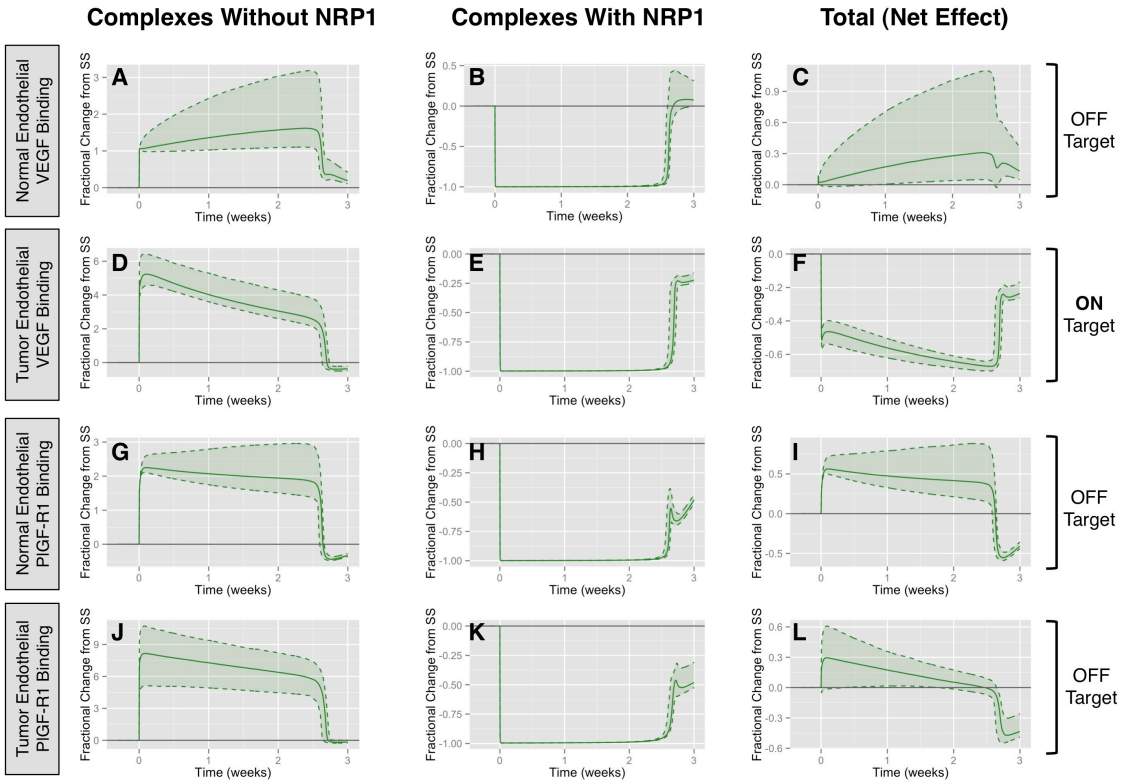

Figure 5-5: Anti-NRP1 response metrics.

A-C, Time course plots of the relative change in aggregate binding of both VEGFA isoforms to VEGFR1 and VEGFR2 in endothelial cells of normal tissues. Solid lines indicate the breast cancer population median at each time point, and dashed lines indicate the 5th and 95th percentiles. VEGF binding to VEGFR1 and VEGFR2 in the absence of NRP1 increased in response to anti-NRP1 (A), whereas the level of VEGF-VEGFR complexes including NRP1 decreased (B). The net effect, which corresponds to the biological signal of interest, was an elevation in VEGF-VEGFR binding. **D-F,** VEGF-VEGFR binding on endothelial cells of the tumor compartment had a net negative change in response to anti-NRP1. **G-L,** Binding of both PlGF isoforms to VEGFR1 followed a similar pattern in the normal and tumor compartments, but both had net elevations. The drug response metrics used in subsequent analyses were the net effects for each complex (C, F, I, L). All results here are from the breast cancer virtual population; the kidney and prostate virtual populations had similar ranges of variability.

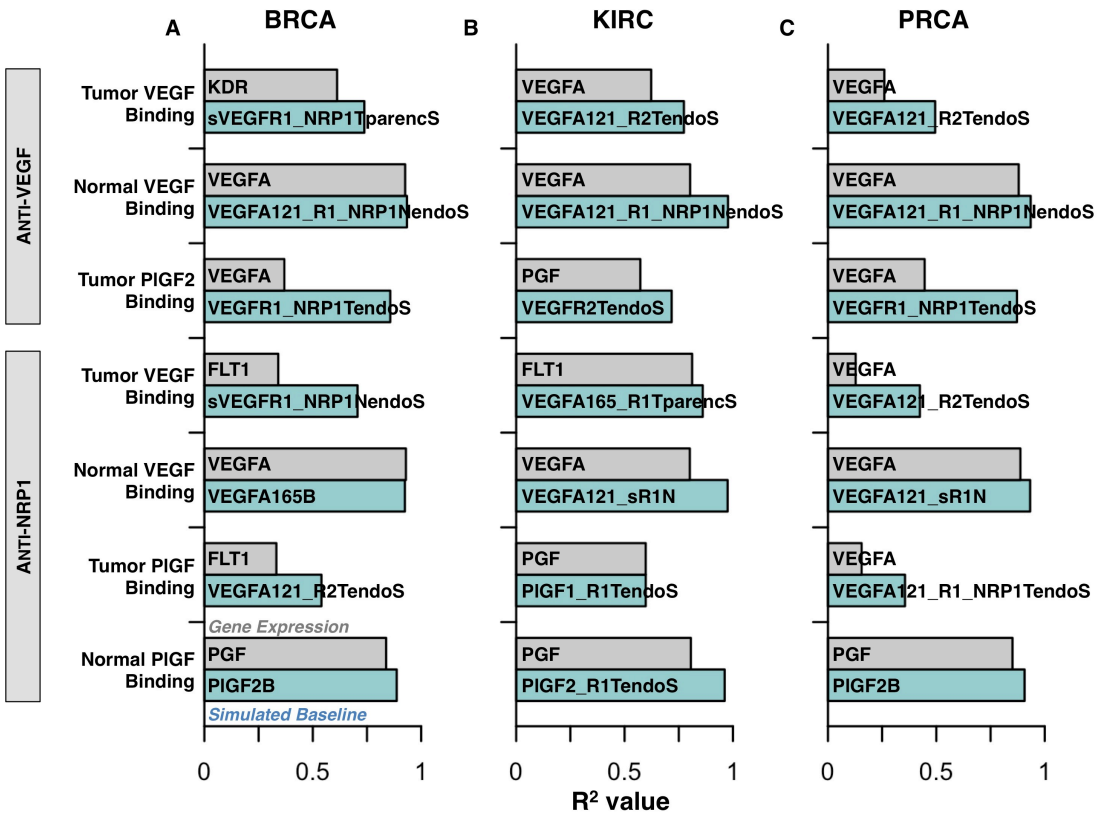

Figure 5-6: Performance of gene expression and simulated baseline as biomarkers of drug response metrics.

A-C, The best single predictor for each metric was almost always a simulated baseline variable. Bars correspond to the highest squared Pearson correlations for each metric in the breast (A), kidney (B), and prostate (C) virtual populations. Grey bars correspond to the best predictor out of the five gene expression variables, with the gene whose correlation was highest listed inside the bar. Blue bars correspond to the best predictor out of the 324 simulated steady-state variables, with the variable listed inside the bar. Simulated variable names ending in "TendoS" indicate that a complex is found in the **T**umor compartment on the **endo**thelial cell **S**urface. An "N" in place of the "T" indicates the **N**ormal compartment. If "parencS" is present in place of "endoS", the variable is located on the **parenc**hymal cell **S**urface. Simulated variable names ending in "B" indicate the ligand is present in the **B**lood.

correlations than the best gene expression variable. This is expected as the computational model incorporates the expression of all the genes; thus even though we used a single simulation variable as a predictor, we actually account for variation in all five genes.

When comparing the best predictors across cancer types, we found that the metrics accounting for ligand binding in the normal compartment consistently had high correlations with the ligands that were part of those complexes (i.e. *VEGFA* for normal compartment VEGF binding in response to both anti-VEGF and anti-NRP1, and *PGF* for normal compartment PlGF binding in response to anti-NRP1). This was not the case for the complexes in the tumor compartments. We also found that plasma ligand concentrations, which are often measured in patients, were not typically good predictors, with the exception of the high correlation between plasma PlGF2 concentration and normal compartment PlGF binding in response to anti-NRP1 in breast and prostate cancers.

5.4 Discussion

Here we studied the baseline VEGF signaling and responses to VEGF-targeting therapies in a virtual population of cancer patients. The virtual populations were built based on actual patient-derived cancer gene expression data and knowledge of the molecular interactions between VEGF ligands and their receptors, as well as transport of VEGF ligands and therapeutic agents throughout the body. This approach gave us insight into the amounts of receptor binding that are currently impossible to measure *in vivo* in the clinic.

Although previous studies have analyzed bevacizumab in whole-body compartment models of VEGF signaling, this study expands on that work in three ways: first, we added PlGF into the model; second, we introduced patient-specific variability in protein production rates using gene expression data; and third, we simulated the effects of an anti-NRP1 antibody. PlGF binding to VEGFR1 increased in response to both therapies, thus the inclusion of PlGF may allow us to account for intrinsic resistance mechanisms due to up-regulated PlGF signaling. The

effects of PlGF signaling have not been studied as extensively as VEGF, but there is some evidence that it promotes angiogenesis [161]; thus, enhancement of PlGF-VEGFR1 signaling is a potentially deleterious off-target effect.

Parameterizing protein production rates based on gene expression data allowed us to create virtual population models. This allowed us to computationally compare drug response profiles in multiple cancer types. Although the range of variation in the drug response metrics tended to be similar in different tumor types, the different gene expression covariance matrices between populations resulted in differing dependencies of drug response metrics on gene expression variables. For example, the best gene expression predictor of bevacizumab-induced reduction in tumor VEGF binding in breast cancer was *KDR*, the gene encoding VEGFR2, whereas in kidney and prostate cancer *VEGFA* was the best predictor. In clinical trials, prospective biomarkers for bevacizumab can succeed in some trials while failing to separate patients in other trials. Our results suggest that differences in gene expression co-variation may be responsible for differing performance of biomarkers between cancer types.

Previous studies involving computational modeling of anti-NRP1 antibodies involved single compartment models [113]; here, our use of multi-compartment models allowed drug pharmacokinetics to be taken into account. We found that the anti-NRP1 antibody had similar effects to bevacizumab: VEGF was redistributed from the tumor compartment to the normal compartment, while some up-regulation of PlGF-VEGFR1 binding occurred. However, there were slight differences, including the reduced range of plasma VEGF up-regulation in the anti-NRP1-treated population. An anti-NRP1 antibody has been tested in the clinic [162]; our results could be useful in suggesting biomarkers that may be able to separate responsive patients from non-responders.

One limitation of this approach is that the only data available for validation at present are plasma measurements of ligand concentrations. Additional measurements will allow for better validation. Gene expression data paired with clinical outcomes for VEGF-targeting drugs will

also be needed to fully validate this approach. However, our results do demonstrate the potential of integrating models describing the kinetics of molecular interactions with patient-specific data.

6 Effect of Cell Heterogeneity on Anti-VEGF Drug Response

6.1 Summary

VEGF-targeting therapies inhibit tumor angiogenesis by preventing VEGF-induced sprouting of new blood vessels from existing vasculature. A difficulty in using gene expression data as a biomarker of the effectiveness of VEGF-targeting therapies is that it is unknown which cells in a tumor contribute to measured gene expression variability. While cancerous tumor cells often comprise the majority of the cells in a tumor, stromal cells such as endothelial cells, fibroblasts, and immune cells may be present in significant quantities as well. Here, we investigated the consequences of assuming that a significant amount of the variability in expression of the genes encoding VEGFR-1, VEGFR-2, and NRP1 arises due to varying production by endothelial cells instead of tumor cells. We use a computational framework previously developed for assessing patient variability in tumor cell expression and adapt it so that production rates by endothelial cells vary linearly with gene expression. We found that drug response metrics maintained the same pattern of association with ligand gene expression, but the receptor expression often had opposite effects when expression variability was on endothelial cells.

6.2 Methods

The computational models used in this chapter were described in Chapter 5. In Chapter 5, all simulations were configured so that variability in the expression of genes encoding receptors led to variability in tumor cell receptors. In this section we test the effect of attributing gene expression variability to the endothelial cells of the tumor compartment instead of the tumor cells. When endothelial cell receptor production rates varied with gene expression, the tumor cell

receptor production rates were held constant at their median values. We considered only the breast cancer dataset in this chapter.

6.3 Results

We first compared how receptor variation on endothelial cells changed the steady-state relationships between gene expression and three key model outputs, relative to the case where receptor variation was primarily on tumor cells. VEGFA expression had the strongest association with free VEGF, endothelial VEGF binding, and parenchymal VEGF binding, regardless of compartment and which cell type had receptor variation (**Figure 6-1**). Receptor variation only had a strong correlation in the tumor compartment (**Figure 6-1A**). The correlation of receptor expression with free VEGF in the tumor compartment was negative regardless of which cell type had receptor variation (top panel). However, the sign of the correlation switched for endothelial and parenchymal (tumor) cell variation (middle and bottom panels, respectively). Receptor gene expression had positive correlations with VEGF binding on the cell type on which receptor variation occurred, and negative correlations on the cell type for which receptor levels were fixed to the median. We examined the correlations of gene expression with each steady-state untreated simulated variable individually as well (**Figure 6-2**). This demonstrated that the trends noted in Figure 6-1 held across individual variables. Outside of the VEGF complexes on tumor and endothelial cells in the tumor compartment (labeled in **Figure 6-2**), there was little difference in Pearson correlation in the population where variation occurred on endothelial cells.

To assess how endothelial cell receptor variation affects drug response, we simulated the addition of an anti-VEGF antibody (bevacizumab) and an anti-NRP1 antibody. We examined the range of variation in tumor endothelial cell VEGF binding and found that whereas the bevacizumab response profile had only a slight increase in the range in variation across the population, the anti-NRP1 response had wider variation (top panels of **Figure 6-3**). When we compared the time-averaged responses in a patient-specific manner, we found a strong correlation

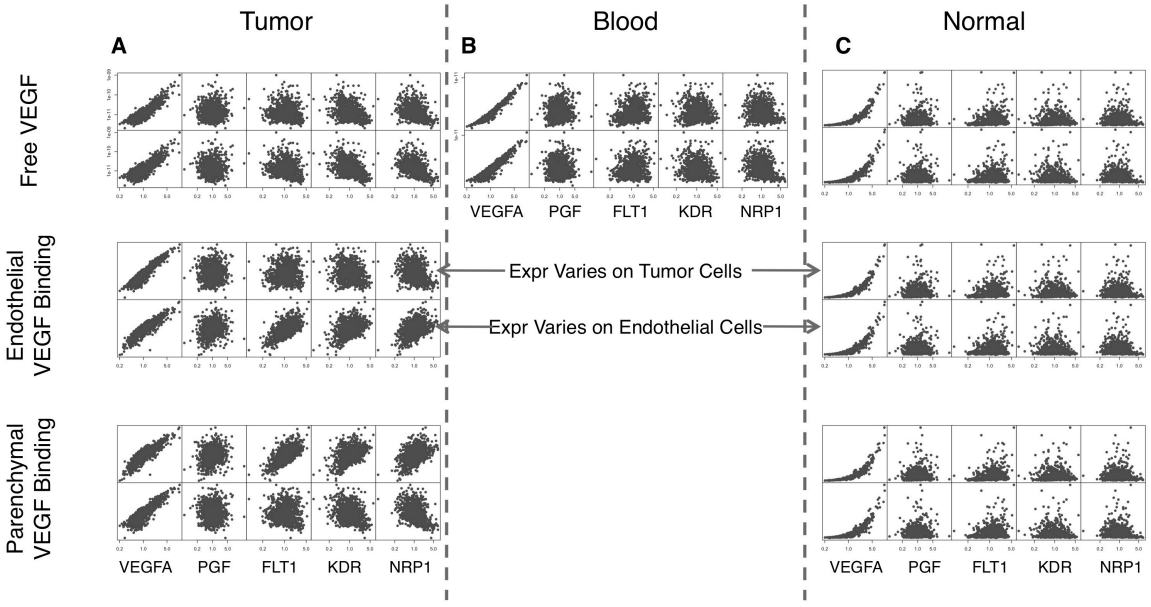

Figure 6-1: Effects of receptor variation cell type on compartment VEGF binding profiles.

A-C, Gene expression versus free $VEGF_{165}$+$VEGF_{121}$ (top), endothelial VEGF binding (middle), and parenchymal VEGF binding (bottom) in the tumor (A), blood (B), and normal (C) compartments. In each rectangle of 10 plots, the top five correspond to simulations where tumor cell receptor levels vary, while the bottom five correspond to simulations where the endothelial cell receptor levels vary. The *x*-axes always correspond to the five gene expression variables.

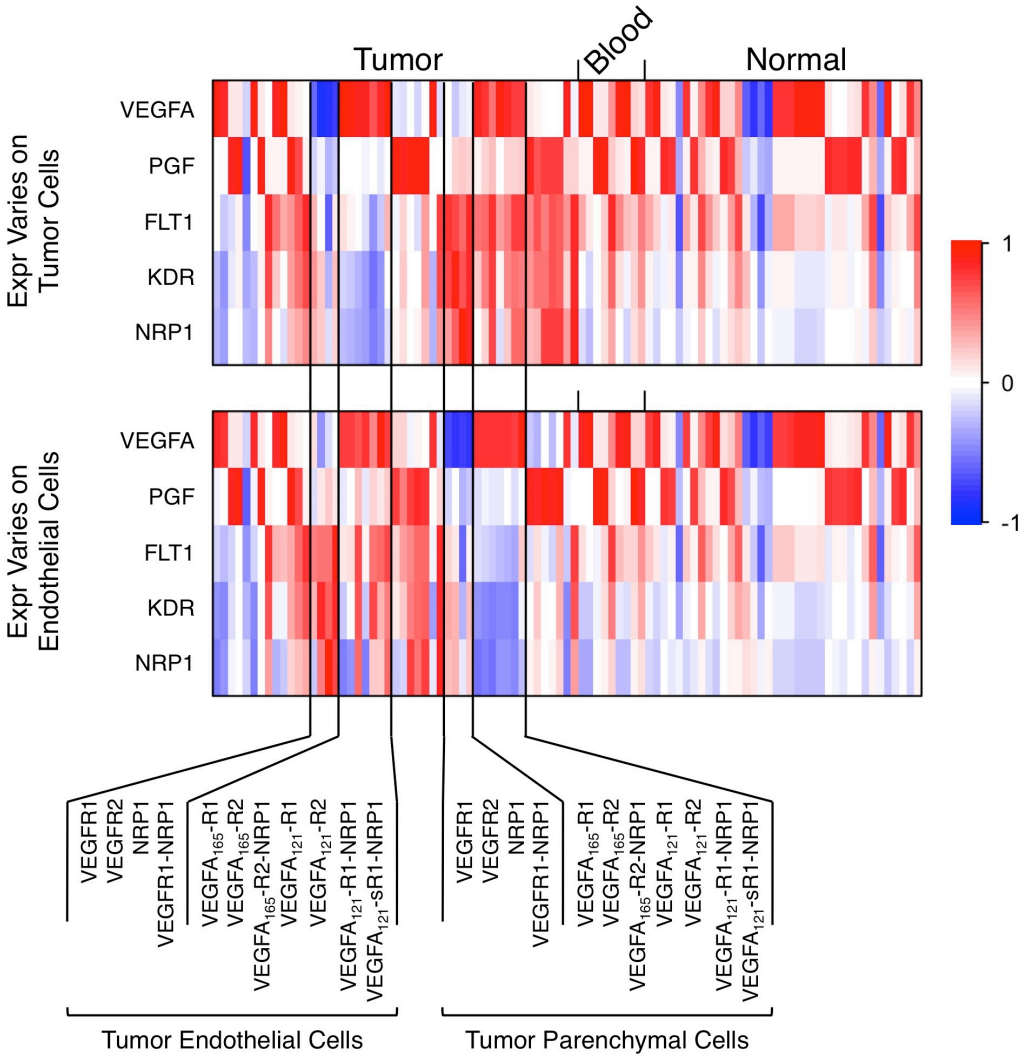

Figure 6-2: Expression-baseline correlation dependence on cell type of receptor variation.

The heatmaps compare the Pearson correlation between gene expression and individual simulated variables, with each row corresponding to the labeled gene and each column corresponding to a different simulation variable. Compartments are noted along the top. The blood compartment has fewer variables than the other two compartments because only soluble complexes are present. The normal compartment has fewer variables than the tumor compartment because some of the normal compartment variables have variances of zero. These are omitted because Pearson correlation coefficients cannot be calculated for these variables. The labeled variables are the ones of interest that display differing patterns depending on the cell type with receptor variability.

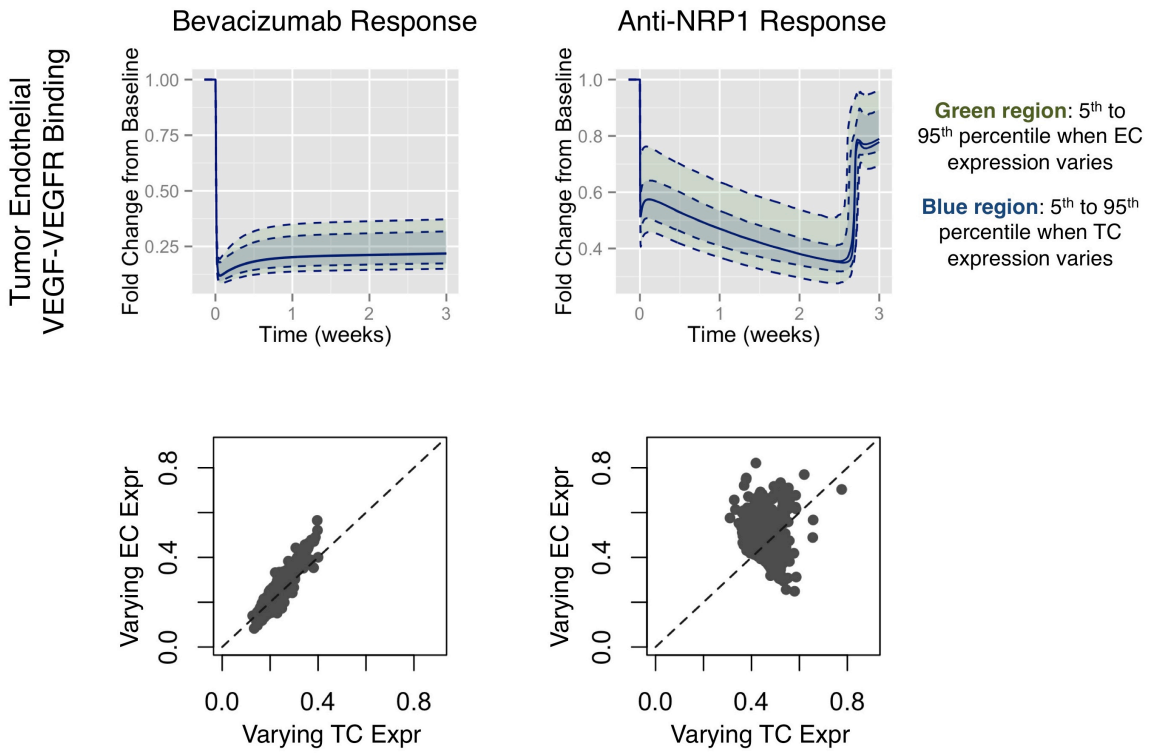

Figure 6-3: Drug response dependence on receptor cell type variation.

The range of variation when receptor levels varied on endothelial cells was higher in response to anti-NRP1 (top right panel) but not in response to anti-VEGF (top left panel). The fold-change from baseline of tumor endothelial VEGF binding was calculated by dividing each time point by the initial untreated level. Blue and green shaded regions indicate the range of population variability when receptor expression varies on tumor and endothelial cells, respectively. The time average of the fold-change from baseline of tumor endothelial binding in response to anti-VEGF is similar when regardless of which cell type has receptor variation; the response to anti-NRP1 does not have this property.

between the populations with receptor variation on endothelial cells and on tumor cells (bottom left panel of **Figure 6-3**). In other words, if patients were ranked from most responsive to least responsive, the order would not change significantly regardless of which cell type contributed cell type receptor variation. This was not the case for the average anti-NRP1 response (bottom right panel of **Figure 6-3**): here, the ranking of patients in the population with varying tumor cell receptor levels would be drastically different than that of the patients in the population with varying endothelial cell receptor levels.

Pearson correlations between gene expression and time-averaged post-bevacizumab simulated variables did not change with differing receptor variation cell types (**Figure 6-4**). Gene expression of the three receptors was highly correlated with the average change in VEGF complexes on both cell types, regardless of which cell type had varying receptor expression. The pattern of correlations between gene expression and time-averaged post-anti-NRP1 simulated variables was less distinct (**Figure 6-5**). Expression of NRP1 had higher correlations with VEGF complexes on endothelial cells in the tumor compartment when receptors varied on endothelial cells compared to when receptors varied on tumor cells. Overall these results suggested that while gene expression-based predictors of bevacizumab response may not be sensitive to assumptions about which cells have varying receptor levels, predictors of anti-NRP1 response are.

6.4 Discussion

A major assumption in building population-based computational models of VEGF signaling involves assigning VEGFR1, VEGFR2, and NRP1 gene expression variability to specific cell types within the tumor compartment. Here we tested how this assumption affects both the untreated baseline levels of VEGF signaling complexes, as well as the drug response profiles. The effects on baseline signaling were somewhat expected: receptor gene expression was strongly correlated with the levels of VEGF complexes on whichever cell type had varying receptor levels. Notably, VEGFA expression was strongly correlated with levels of baseline VEGF complexes in

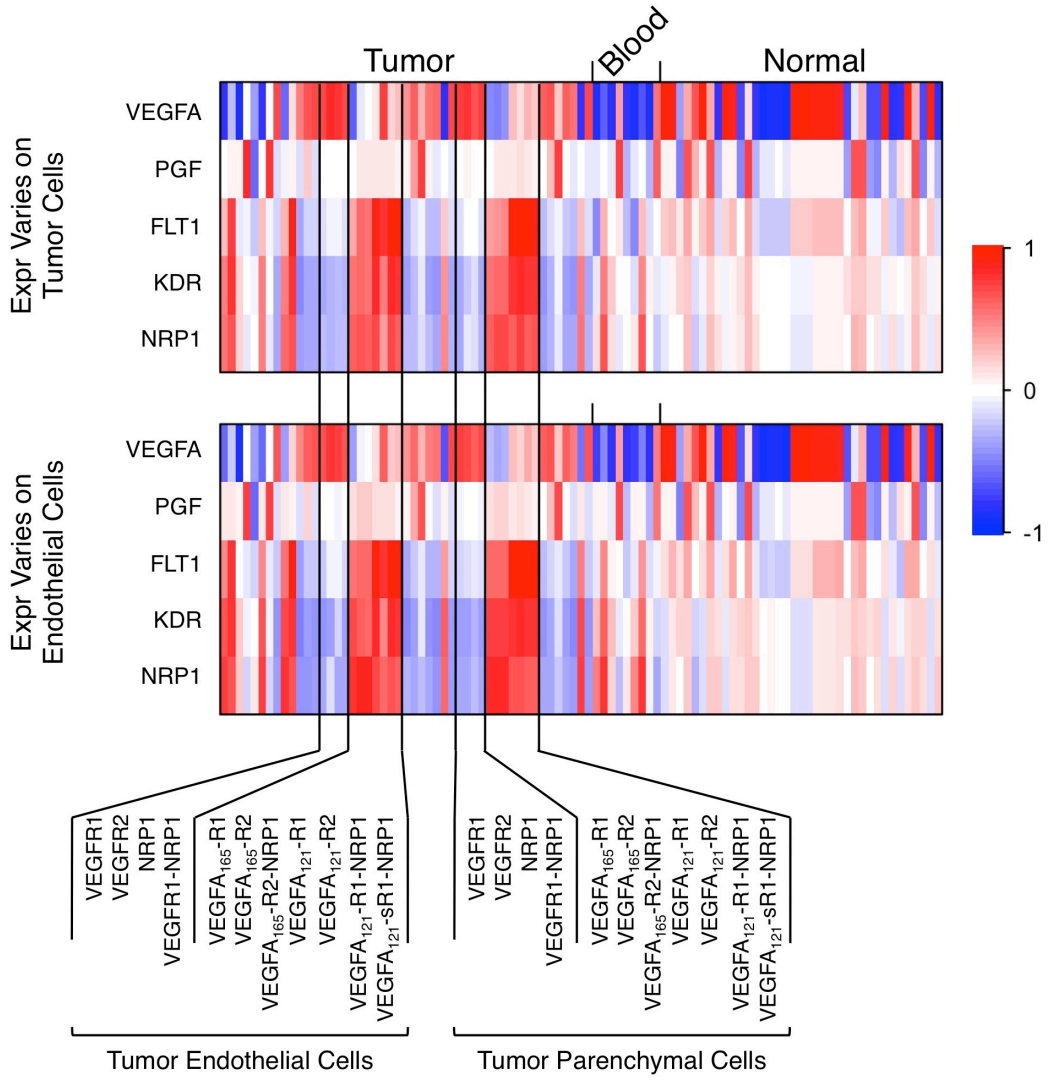

Figure 6-4: Effect of receptor variation cell type on expression-bevacizumab response correlation.

Analogous to Figure 6-2, but the simulated variables along the columns are now the time-averaged fold-change in post-bevacizumab values. The color bar on the right displays the scale, with red indicating strong positive Pearson correlations and blue indicating strong negative Pearson correlations.

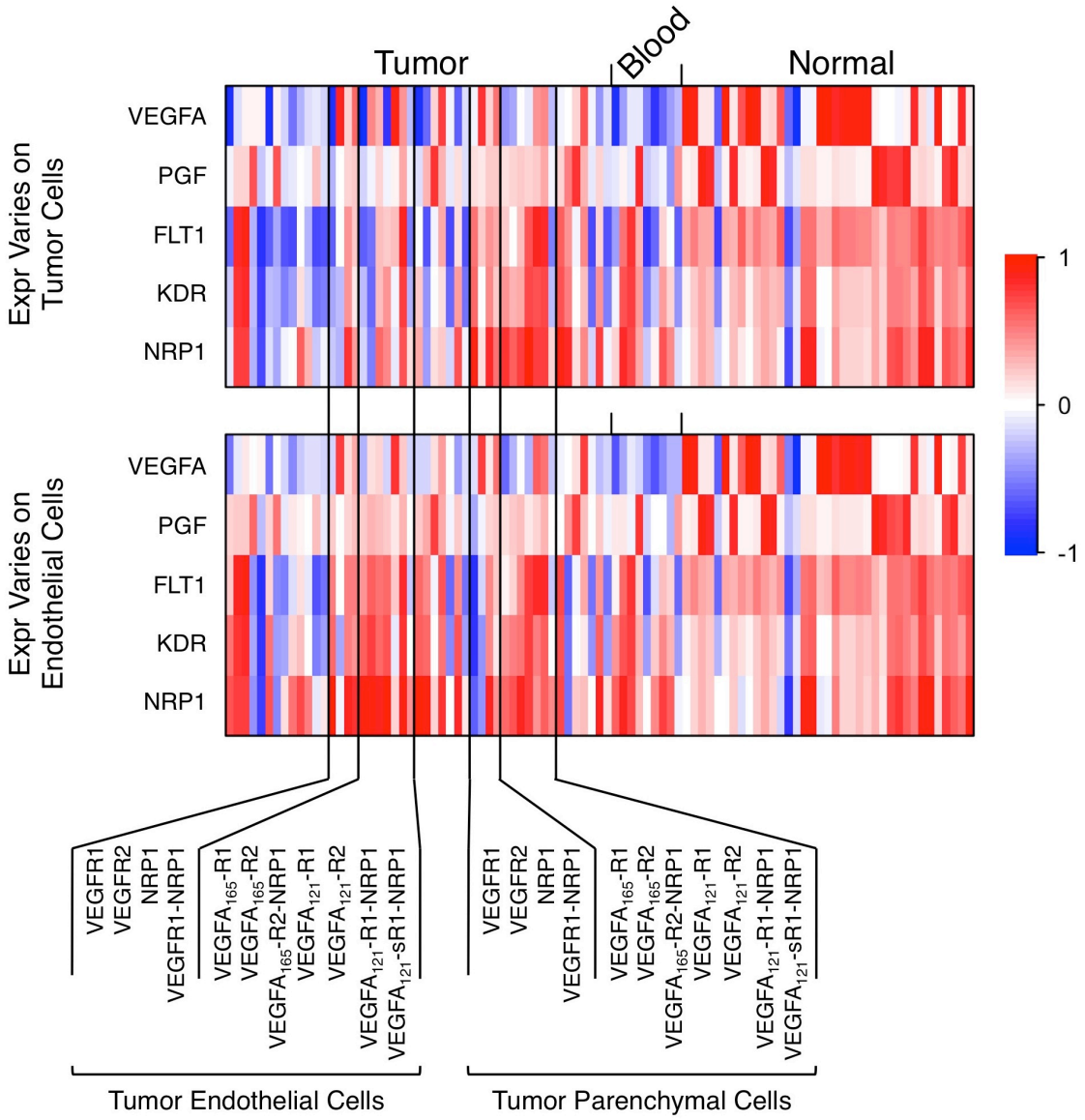

Figure 6-5: Effect of receptor variation cell type on expression-anti-NRP1 response correlation.

Analogous to Figure 6-2, but the simulated variables along the columns are now the time-averaged fold-change in post-anti-NRP1 values. The color bar on the right displays the scale, with red indicating strong positive Pearson correlations and blue indicating strong negative Pearson correlations.

both populations. This suggests that simulated predictions of the effect of VEGFA expression on VEGF signaling are insensitive to the assumption of which cell types contribute gene expression variability.

When considering the impact of the receptor variability cell type on drug response, bevacizumab response did not depend on the cell type, but anti-NRP1 response did. This implies that developing biomarkers for response to bevacizumab based on this approach may not require further knowledge as to which cell types in a tumor contribute gene expression variability. However, cell-type specific measurements of receptor gene expression would be helpful for developing anti-NRP1 biomarkers. This discrepancy could be due to the fact that bevacizumab binds to a ligand while anti-NRP1 binds to a receptor. This could have implications for other VEGF receptor-targeting therapies, such as the anti-VEGFR2 antibody ramucirumab.

Here, we tested two extremes: receptors either varied solely on tumor cells or solely on endothelial cells. A more realistic simulation would include variation on both cell types. Measurements of cell-specific receptor gene expression will be useful in further fine-tuning the expression-based computational models of patient populations. Pairing these measurements with relative frequencies of cell types in a patient's tumor would allow us to simulate varying levels of receptors on both cell types.

7 Personalized Computational Models of EGFR/ErbB Family Signaling Reveal Mechanisms of Intrinsic Resistance in Patient Subgroups

7.1 Summary

A major obstacle in oncology drug development is the variability among patients of effectiveness of therapies, due to the heterogeneity of tumors across populations. Molecular biomarkers such as the expression levels of genes or proteins have proven useful, particularly in the epidermal growth factor (EGF) family, but these kinds of markers are still unable to accurately predict all responsive patients. These biomarkers are typically univariate and linear, whereas multivariate, nonlinear biomarkers are needed to adequately describe the network of molecular interactions targeted by a drug. Here, we present a method for building multivariate biomarkers that incorporates BOTH patient-specific transcriptomic/proteomic data AND detailed mechanistic models of the nonlinear interactions between ligands and receptors. We applied the method to the EGFR/ErbB family and showed that the personalized models accurately captured variability in receptor phosphorylation. Before the addition of drugs, the models behaved in a relatively monotonic fashion, with signaling outputs following closely the expression of the key ligands. However, the response to the addition of drugs was much more complex. We simulated the addition of three antibody drugs that each target one of EGFR, HER2, and HER3 and applied principal component analysis to characterize the post-therapy time course response. We derived metrics that accounted for both target-specific and off-target effects. We then used multivariate supervised learning methods to develop predictive biomarkers. We found that biomarkers derived from gene expression data were outperformed by biomarkers derived from simulated baseline tumor behavior (i.e. that combined quantitative mechanistic information with gene expression

data). This suggested that linear transformations of transcriptomic and proteomic data may not be adequate for predicting drug response; instead, the nonlinear transformation central to the computational model is needed. With paired transcriptome/proteome and clinical drug response data, the method proposed here could yield improvements in predictive biomarker development.

7.2 Introduction

Drugs that target epidermal growth factor (EGF) family signaling have shown success as cancer treatments. Several of these target the receptors for the EGF family – the EGFR/ErbB/HER family of receptors (**Figure 7-1B**). The EGFR-targeting antibody cetuximab [163] and the HER2/ErbB2-targeting antibodies trastuzumab [56] and pertuzumab [57] are approved for use in certain cancer types. A HER3/ErbB3-targeting antibody, MM-121, has shown promise in preclinical models [164] and early-stage clinical trials. Tyrosine kinase inhibitors that repress receptor phosphorylation, such as erlotinib [165] and lapatinib [166], have also been approved. A major problem with these and other cancer therapies is that the tumors of many individuals have intrinsic resistance and do not respond. This is likely due to the tumor-to-tumor variation in expression not only of the targets but also of the other proteins in the protein interaction network of the target. Predictive biomarkers are needed that can distinguish between responders and non-responders so that these costly therapies can be selectively administered to patients who will benefit.

Predictive biomarkers already exist for some therapies, with one of the best known examples being HER2 amplification predicting response to trastuzumab, a HER2 antibody. Testing for HER2 overexpression/amplification by immunohistochemistry and in situ hybridization can separate patients into two to four groups, with increasing rates of pathologic complete response in groups with higher HER2 amplification [167]. However, even among patients with the highest levels of HER2, the drug is not completely effective: at a median of 3.9 years of follow-up in two large clinical trials, 85.8% of treated patients were disease-free

compared to 75.8% of controls [168]. Thus, 14.2% of HER2-positive patients relapse despite treatment. In colorectal cancer, expression of two EGFR ligands is predictive of cetuximab response, as is mutation status of the downstream signaling protein K-ras [169]. Studies of mechanisms of resistance in cell line and xenograft models can be useful in the development of predictive biomarkers for the clinic. For example, ErbB2 amplification and overexpression of the ErbB3 ligand heregulin (HRG) were demonstrated to promote cetuximab resistance in cell lines [170]. HRG expression was demonstrated to be necessary for responsiveness to MM-121 in cell lines and xenografts [164]. HER2 amplification by itself may not always be sufficient to predict response to trastuzumab: the oligomeric state of HER2 is also important for prediction of response in cell lines, with high levels of homodimers promoting sensitivity and high levels of ligand-induced heterodimers promoting resistance [171].

Here, we present an approach for integrating genomic data from patient tumor samples into computational models of interactions between ligands and receptors in the EGF family. We build upon work previously published by other groups: we start with the model of Shankaran *et al.* [172], which was used to determine patterns of receptor dimerization that lead to phosphorylation. To this model, we add an additional ligand and adapt the geometry so that it represents a human breast tumor. We also add three EGF family-targeting drugs, cetuximab, pertuzumab, and MM-121, based on the kinetics and interaction topology previously published by Schoeberl *et al.* [173]. We calibrate a "mean" model based on published ligand concentrations in cancer patients, and then parameterize the protein production rates using individualized data. Thus, we generate a population of models, each parameterized to a different individual. This allows us to simulate the drug responses of the virtual patient population **(Figure 7-1A)**. From this data, we use several machine learning approaches to develop predictive biomarkers for drug response, and show that these biomarkers, mechanism-based nonlinear latent variables that could not be obtained from expression data alone, perform better than biomarkers derived only from gene expression data.

7.3 Methods

7.3.1 Mathematical model of the EGFR family

To predict the EGF family ligand-receptor binding based on patient gene expression data, we used the mathematical model developed and published by Shankaran *et al.* [172]. This model was comprised of the binding and trafficking reactions involving EGF, HRG, EGFR, HER2/ErbB2, and HER3/ErbB3 (**Figure 7-1B**) in cultured cells. We omitted the late endosome compartment because molecular species in the late endosome did not contribute to the dynamics of other species and did not contribute to the phosphorylation measurements. This meant that molecules exiting the early endosome were either recycled or degraded as opposed to entering the late endosome for subsequent degradation.

We adapted the model so that it represented a tumor while retaining the binding and trafficking kinetics that had previously been validated; this required making changes to the geometry of the model but not to the molecular interactions. Shankaran *et al.* modeled cells with a media volume of $5x10^{-10}$ liters/cell and endosome volume of $3x10^{-14}$ liters/cell. We modeled tumor tissue consisting of tumor cells and interstitial space. The interstitial volume fraction was 0.58 cm^3/cm^3 tissue, while the tumor surface area-to-volume ratio was 1534 cm^2/cm^3 tissue. Assuming a surface area of 1000 um^2/cell, this equates to $1.534x10^8$ cells/cm^3 tissue. We use this cell density to convert the endosome volume to a tissue volume basis, yielding an endosome volume fraction of 0.0046 cm^3/cm^3 tissue. We assumed that the surface area of endosome was equivalent to the surface area of the cell membrane, giving the endosome the same surface area-to-volume ratio of 1534 cm^2/cm^3 tissue. We simulated the addition of 10 ng/mL EGF and 40 ng/mL HRG into the interstitial space to confirm that our model yielded similar results to those of Figure 3 in Shankaran *et al.* (**Figure 7-2**).

To generate patient-specific models of EGF family signaling, we used a unique set of protein production rates in each model. To determine these rates, we first determined production

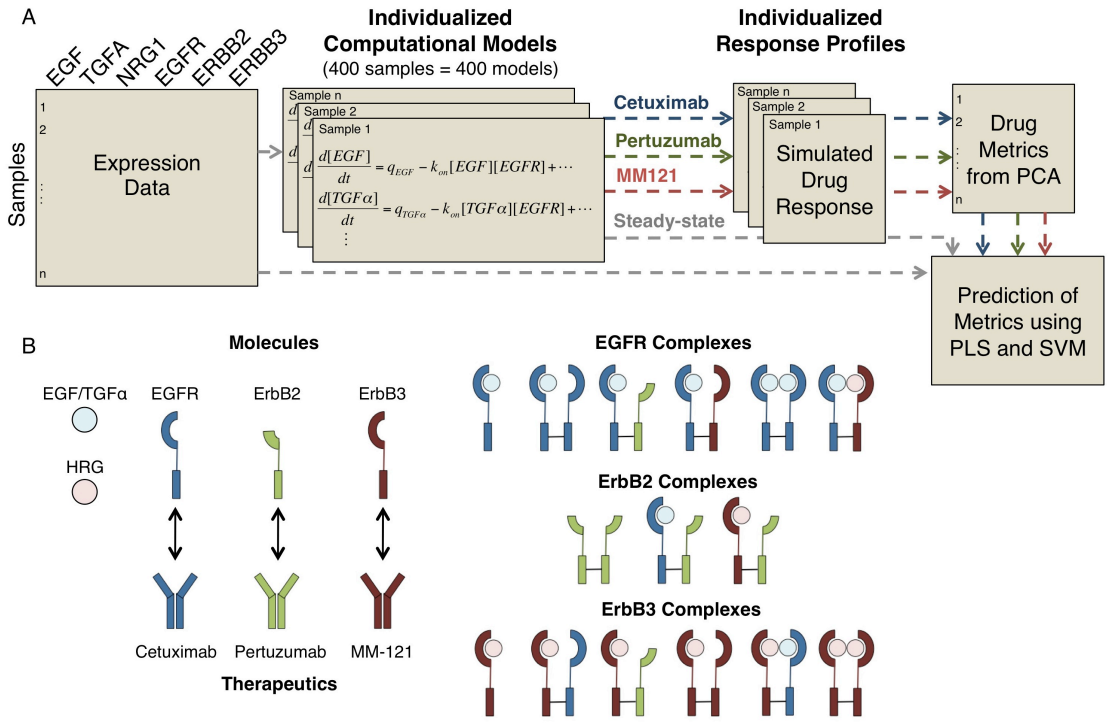

Figure 7-1: Overview of computational approach for identifying predictive biomarkers.

A, Gene expression data for six EGF family genes were used to simulate the ligand-receptor interactions in a population of patients. Time course data describing the response of the simulated population to three therapies was analyzed to find potential response metrics. These metrics were then modeled using supervised learning to compare the performance of transcriptomic/proteomic data with simulated baseline predictions. **B,** The molecules in the computational model include the ligands EGF, TGFα, and HRG, the receptors EGFR, ErbB2, and ErbB3, and the therapeutic monoclonal antibodies cetuximab, pertuzumab, and MM-121. All free receptors and ligand-receptor complexes can be found at the surface of the cell as well as in the endosome.

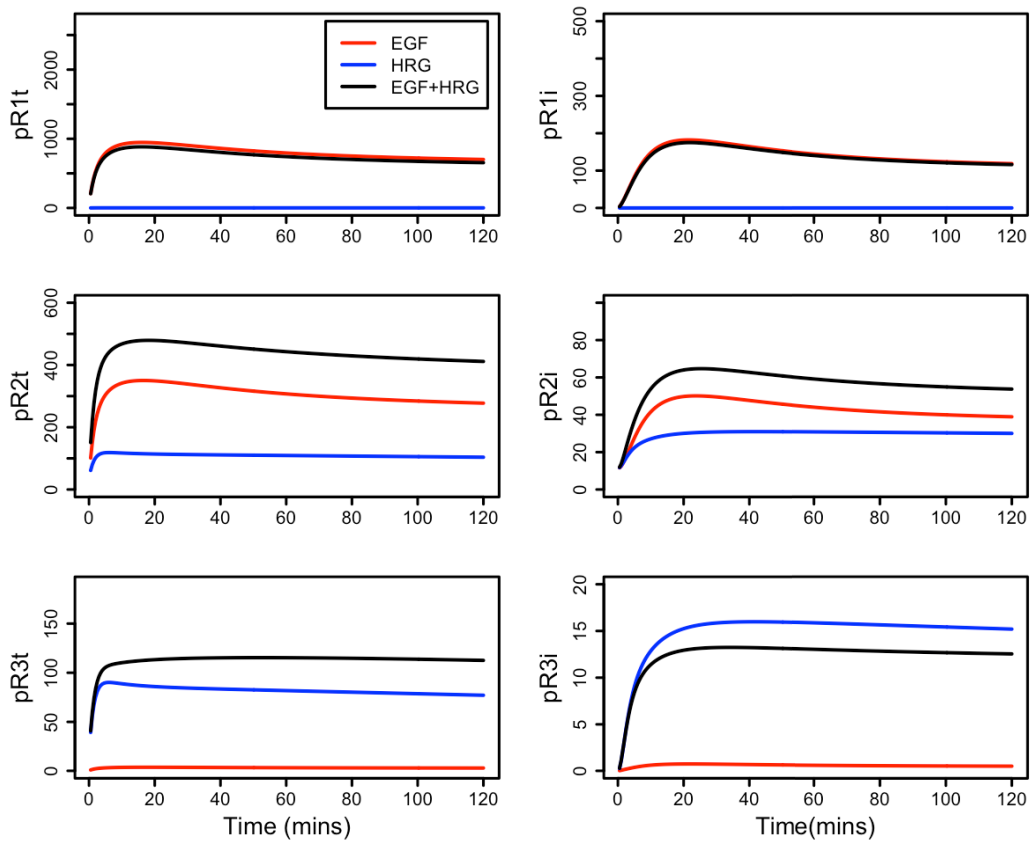

Figure 7-2: Comparison of simulation results to Shankaran *et al.*

Our implementation of the model returned identical results to those in Figure 3 of Shankaran *et al.*, which shows the response of the D20 clone of mammary epithelial cells (85,896 EGFR/cell, 643,777 ErbB2/cell, and 28,770 ErbB3/cell) to 12 ng/mL EGF alone, 40 ng/mL HRG alone, and EGF and HRG combined for two hours.

rates that yielded protein levels comparable to the median protein level measurements from populations of cancer patients. We then varied these median production rates in each patient based on expression data from the Cancer Genome Atlas (TCGA).

We also added three therapies into the model based on the work of Schoberl *et al.* [173]: cetuximab, pertuzumab, and MM-121. These drugs target EGFR, ErbB2, and ErbB3, respectively. Each antibody binds to its target receptor and prevents ligand binding (for EGFR and ErbB3) and receptor dimerization (for all three receptors). The antibodies are able to bind two receptors, so ternary complexes consisting of an antibody and two receptors are included in the model. Internalization of antibody-bound receptors does not occur except in the case of MM-121; here, the ternary complex consisting of MM-121 and two ErbB3 receptors can enter the early endosome, where it is recycled or degraded. We included pertuzumab as an anti-ErbB2 antibody instead of trastuzumab because the mechanism of action of trastuzumab is less clear and may involve antibody-dependent cell-mediated cytotoxicity, whereas pertuzumab acts by blocking dimerization [174].

The mathematical models of the EGF family interactions were coded in Fortran and integrated using a fourth-order Rosenbrock step [175], implemented using the Intel *ifort* compiler.

7.3.2 Gene and protein expression datasets

We used TCGA data (cancergenome.nih.gov) to determine protein production rates in the model. We used RNA-Seq measurements of *EGF*, *TGFA*, and *NRG1* expression to determine the EGF, TGFα, and HRG production, respectively. We used reverse phase protein array (RPPA) measurements of EGFR, ErbB2, and ErbB3 levels to estimate the steady-state levels of these proteins, and from this their production rates. We also used RPPA measurements of receptor phosphorylation for comparison with steady-state model predictions. In total, there were 400 patients with both RNA-Seq and RPPA data.

7.3.3 Outputs of the model

We simulated the population several times: once to determine the steady-state baseline levels of the molecular complexes in all 400 individuals across the population, and once for each of the three drugs at multiple concentrations. For each of the three drugs, each individual in the population started from the corresponding steady-state baseline for that individual. The observation time was 48 hours following treatment. The dose of each drug was chosen by performing dose response simulations across the population and choosing the minimum dose that was at least moderately effective in most patients. The individual simulated baseline molecular complex levels were subsequently used as predictor variables in the supervised learning models, while drug response time course-derived variables were used as outcome metrics (see Supervised Learning of Multivariate Predictive Biomarkers below).

7.3.4 Principal component analysis

We performed PCA using the *prcomp* function in the *stats* package in R. We selected the number of components to include in the PCA model so that subsequent addition of components did not significantly increase the total variance explained. We performed rotation of the loadings for these components using the *varimax* function in the *stats* package in R.

7.3.5 Supervised learning of multivariate predictive biomarkers

We used two supervised learning algorithms to create statistical models of the PCA-derived metrics from two sets of predictor variables: the gene/protein expression data and the simulated steady-state model results. For partial least squares (PLS) regression, we used the *plsr* function in the *pls* package in R. The number of components was chosen by cross-validation. For support vector machine (SVM) regression, we used the *svm* function in the *e1071* package in R. We used the default kernel, a radial basis function, with the default value of γ.

We chose parameter values as needed in the supervised learning models by performing leave-one-out cross-validation (LOOCV). Each sample was withheld once. Output values were

predicted for the held out sample using a model trained with all remaining samples. The model goodness of fit was then assessed by finding the R^2 value between the values predicted for the held out samples and the actual data.

7.4 Results

7.4.1 Model expansion and calibration

As a first step toward building patient-specific computational models of EGF family signaling, we began with a previously-published model. Extensive work has been done modeling the interactions of EGF family proteins and downstream signaling [173,176,177]; we selected the recently published model of Shankaran *et al.* [172] as it focused on the relative signaling potency of the various receptor homo- and hetero-dimers. Transforming growth factor-α (TGFα) is an EGFR ligand with potential importance in breast cancer due to its up-regulation in the triple-negative subtype (**Figure 7-3**), but to our knowledge no detailed mathematical models have included TGFα. We therefore expanded the Shankaran *et al.* model to include TGFα based on published data describing the affinities and trafficking of multiple EGFR ligands [178]. The TGFα-EGFR affinity was similar to that of EGF-EGFR, whereas the rate of recycling of TGFα-EGFR complexes was higher than for EGF-EGFR complexes. We therefore modeled TGFα as having the same binding kinetics as EGF and the same recycling rate constant as HRG-ErbB3 complexes, which recycle faster than EGF-EGFR complexes.

The Shankaran *et al.* model was developed for application to cells in culture; to adapt this model so that it represented a tumor, we added production terms for the three ligands and adjusted the geometric parameters, as described in the Methods. We set the receptor production rates to values that yielded densities of 120,000 EGFR/cell, 200,000 ErbB2/cell, and 20,000 ErbB3/cell in the absence of ligands; these values fall in the middle of the range of measured cell densities across cell lines in [172]. We then adjusted the production rates of EGF, TGFα, and HRG to target extracellular concentrations of 158 pM [179], 29 pM [180], and 216 pM [170],

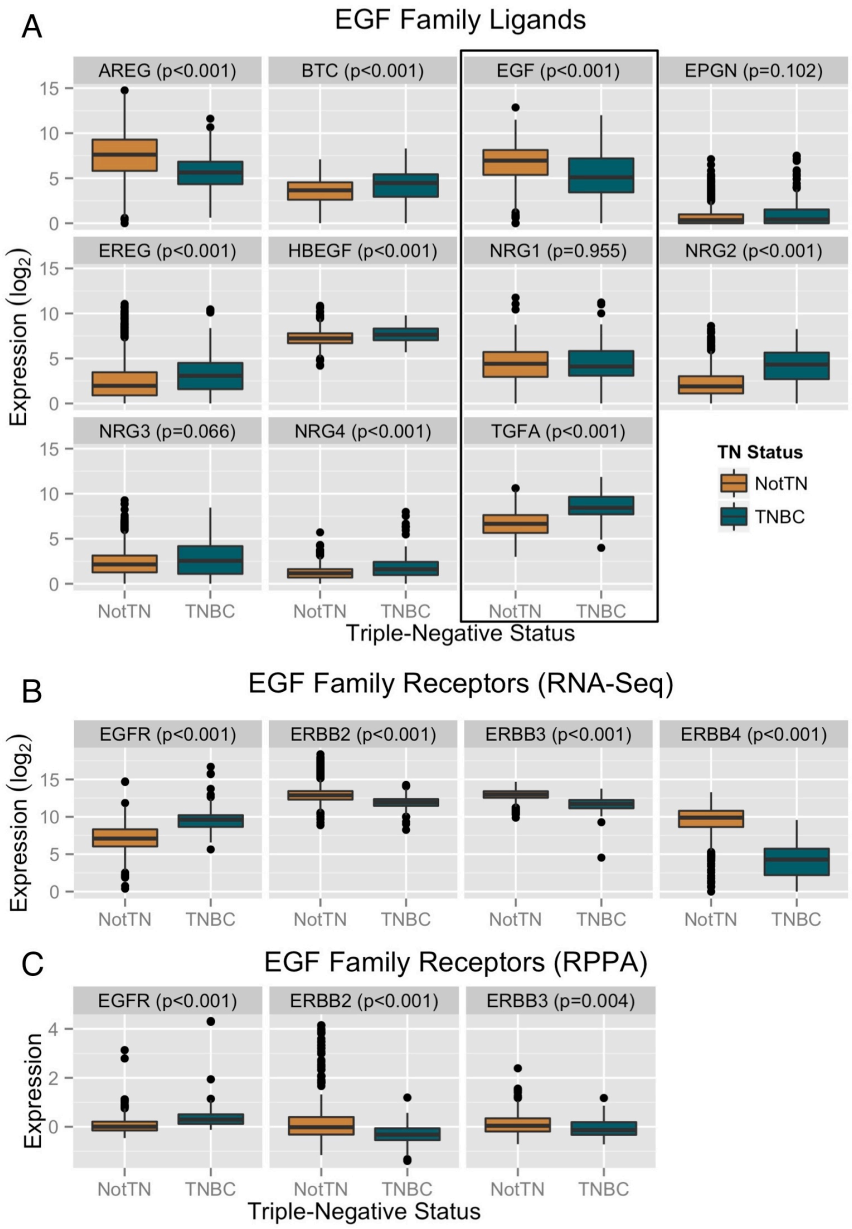

Figure 7-3: Comparison of EGF familiy proteins between triple-negative breast cancer (TNBC) and other breast cancers.

A, Most ligands were significantly altered in TNBC. The ligands included in the computational model in this study are enclosed in the box. **B,** All 4 EGF family receptors were significantly altered in TNBC. **C,** The three receptors with RPPA data were significantly altered. The pattern of receptor alterations agreed between the RNA-Seq and RPPA measurements. *P*-values were calculated using the Mann-Whitney U-test.

respectively. These six production rates served as population median values in the patient-specific models. For a given protein in a particular patient, the protein production rate was determined by dividing the measured expression of that protein by the median expression across the population. This normalized value was then multiplied by the median production rate to yield the individual production rate. This resulted in patient-specific models of EGF family signaling.

7.4.2 Predictions of baseline pre-drug tumor state correlate strongly with gene expression and breast cancer subtypes

The steady-state levels of each of the molecular complexes in the model, simulated for each individual in the population of 400 breast cancer patients, were predicted; these correspond to the baseline, untreated levels for these complexes in tumors. Analysis of the steady-state predictions across the 400 patients revealed extensive correlations among the 50 distinct molecular entities in the model. We used principal component analysis to investigate the correlation structure in depth. We found that the first six principal components contained 97.6% of the overall variance. After rotating the first six component loadings using the varimax procedure, we found that the first four components were significantly associated with the production rate of four of the proteins (**Figure 7-4A** and **7-4B**, components are labeled with the "SS" subscript to differentiate these principal components from later applications of PCA). Specifically, only EGF-containing complexes had high magnitude $PC1_{SS}$ loadings. The same pattern was observed for HRG-containing complexes and $PC2_{SS}$, TGFα-containing complexes and $PC3_{SS}$, and ErbB2-containing complexes and $PC4_{SS}$. The two molecules with high magnitude $PC1_{SS}$ and $PC2_{SS}$ loadings were the surface and endosome EGF-HRG-EGFR-ErbB3 complexes, and the two molecules with high magnitude $PC3_{SS}$ and $PC4_{SS}$ loadings were the surface and endosome TGFα-EGFR-ErbB2 complexes. Thus nearly all of the variation in predicted baseline EGF family signaling could be described by four reduced variables: EGF binding, HRG binding, TGFα binding, and ErbB2 dimerization.

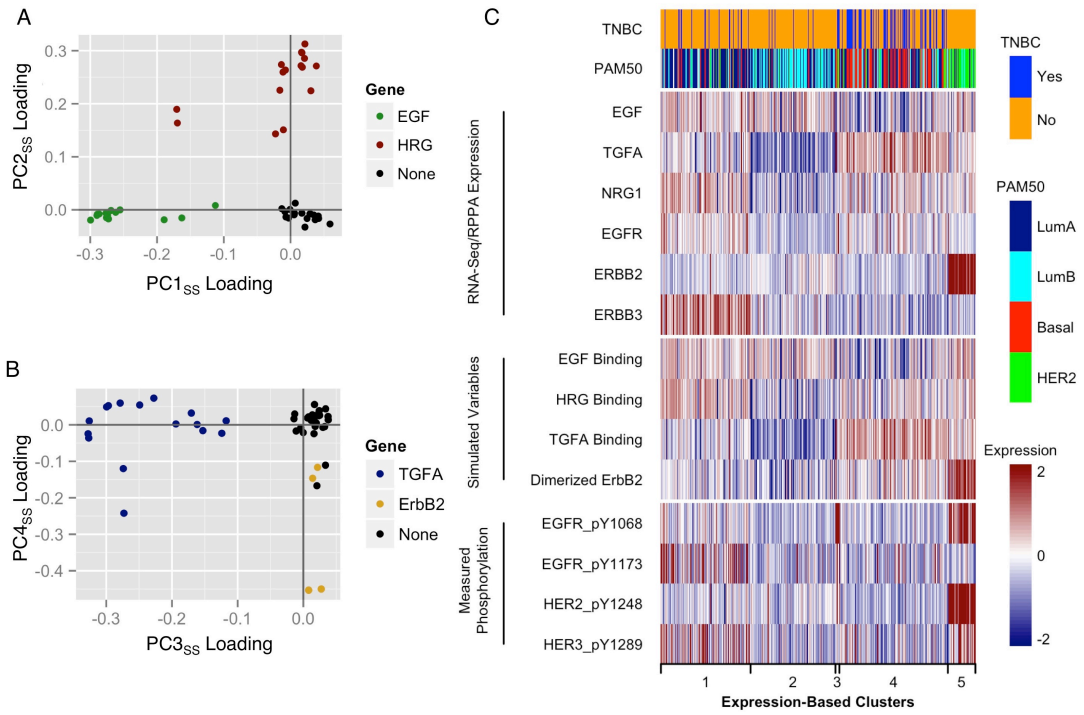

Figure 7-4: Simulated ligand binding profiles are predictive of receptor phosphorylation.

A-B, The rotated principal component loadings for the first 6 PCs (97.6% of the total variance) show ligand-driven patterns for the first three PCs (EGF, HRG, and TGF-α-containing complexes shown in green, red, and blue, respectively), and a receptor-driven pattern for $PC4_{SS}$ (ErbB2-containing complexes shown in yellow). **C,** TN status and PAM50 subtypes (top block), reduced simulated variables (third block), and phosphorylation measurements (bottom block) were associated with gene expression (second block). Samples were ordered according to *K*-means clustering of the samples based on the expression of the six input genes, with the cluster membership labeled at the bottom of the heatmap. Expression, simulated variables, and phosphorylation measurements were all log-transformed if needed, and scaled to have zero mean and unit variance so that the same color scale could be applied to all three data types.

We compared the four reduced simulated variables to previously defined breast cancer subtypes, sometimes referred to as the PAM50 subtypes. When grouping the samples into five gene expression-based *K*-means clusters, we noted that the four PAM50 subtypes aligned well with these clusters, suggesting that EGF family signaling varies in a subtype-specific manner (**Figure 7-4C**, the fifth *K*-means cluster was very small and cannot be seen here). Specifically, the luminal A subtype was strongly associated with overexpression of *NRG1* (HRG) and *ERBB3* (cluster 1), the basal/triple-negative subtype was associated with *TGFA* over-expression (cluster 3), and the HER2-enriched subtype was associated with *ERBB2* over-expression (cluster 4). The luminal B subtype had no over-expression of any of these genes; the only distinctions noted were reduced expression of *TGFA*, *NRG1*, and *EGFR* (cluster 2). Note that while the subtypes tended to cluster together there were some tumors in each subtype that were more similar to those in other clusters (data not shown).

The phosphorylation measurements of these receptors in the TCGA dataset were consistent with model predictions (**Figure 7-4C**, bottom block). Phosphorylation of HER2 at Y1248 was associated with high *ERBB2* expression/predicted ErbB2 dimerization (Pearson correlation $r = 0.82$ for expression, $r = 0.44$ for predicted dimerization – see **Table 7-1**), and phosphorylation of HER3 at Y1289 was associated with *NRG1* expression/predicted HRG binding ($r = 0.19$ for expression, $r = 0.15$ for simulated binding). None of the phosphorylation measurements were associated with the basal/triple-negative subtype, where TGFα binding was predicted to be high. The association of EGFR phosphorylation at two sites, Y1068 and Y1173, appeared to be associated with HER2 and HER3 phosphorylation, respectively (EGFR pY1068-HER2 pY1248 $r = 0.63$ and EGFR pY1173-HER3 pY1289 $r = 0.42$).

Table 7-1: Correlations between gene expression, simulation steady-state, and RPPA phosphorylation measurements.

	Gene Expression Measurements						*Simulated Variables*				*RPPA Phosphorylation*		
	EGF	TGFA	NRG1	EGFR	HER2	HER3	Bound EGF	Bound HRG	Bound TGFA	Dimzd ErbB2	EGFR pY1068	EGFR pY1173	HER2 pY1248
EGF													
TGFA	-0.12												
NRG1	0.12	0.12											
EGFR	-0.07	0.22	0.11										
HER2	0.05	-0.07	-0.03	-0.13									
HER3	0.03	-0.10	0.07	0.05	-0.03								
Bound EGF	**0.99**	-0.15	0.14	-0.07	0.13	0.03							
Bound HRG	0.22	0.10	**0.96**	0.04	0.02	0.11	0.23						
Bound TGFA	**-0.46**	**0.89**	0.09	0.22	0.03	-0.11	**-0.44**	0.03					
Dimerized ErbB2	0.31	**0.53**	0.05	0.04	**0.55**	-0.06	0.32	0.11	0.38				
EGFR pY1068	0.03	0.07	0.03	0.38	0.38	-0.07	0.06	0.03	0.11	0.23			
EGFR pY1173	0.06	-0.02	0.08	**0.57**	-0.21	0.28	0.06	0.05	-0.05	-0.15	0.18		
HER2 pY1248	0.02	0.00	0.04	0.03	**0.82**	-0.03	0.09	0.08	0.09	**0.44**	**0.63**	-0.03	
HER3 pY1289	0.04	-0.07	0.19	0.21	0.23	0.17	0.08	0.15	-0.04	0.06	0.34	**0.42**	**0.43**

7.4.3 PCA of dynamic responses reveals potential mechanisms of resistance

Based on the steady-state model analysis, we expected that groups of patients sensitive to EGF family-targeting drugs would overlap with PAM50 subtypes (i.e. basal/triple-negative patients would respond to the EGFR-targeting drug cetuximab since TGFα binding is elevated in that group). To assess the effects of therapies, we simulated the addition of three monoclonal antibody drugs: cetuximab (anti-EGFR), pertuzumab (anti-ErbB2), and MM-121 (anti-ErbB3). We selected doses by simulating a range of doses in the virtual population and then selecting the dose that yielded a significant amount of population variability (**Figure 7-5A** through **7-5D**). As expected, across the population of patients, cetuximab lowered EGF and TGF-α binding across the patients (**Figure 7-5A**), pertuzumab lowered ErbB2 dimerization (**Figure 7-5C**), and MM-121 lowered HRG binding (**Figure 7-5D**). At high enough doses, targeted binding and dimerization were completely inhibited. Off-target effects were also noted. These were largest for cetuximab, which up-regulated HRG binding (**Figure 7-5B**). Small off-target effects were noted for pertuzumab, in the form of slight down-regulation of EGF, TGFα, and HRG binding. MM-121 appeared to be a drug with high specificity: no off-target effects were observed across the range of doses (data not shown).

At drug doses that yielded a wide range of population variability, we analyzed the time course behavior. There was significant inter-individual variability across the population in terms of the time-course of response to the three drugs (**Figures 7-6, 7-7, and 7-8** for cetuximab, pertuzumab, and MM-121, respectively**)**. The drug effects did not always persist for the duration of the treatment: on-target effects for all three drugs were repressed the most at early time points, followed by a gradual rise throughout the 48-hour treatment (**Figure 7-5E, F, G**). In some patients, cetuximab-induced EGF binding and MM-121-induced HRG binding were actually higher than the baseline level at later time points. Off-target effects, when present, tended to last for the entire treatment duration (**Figure 7-5H, I, J**).

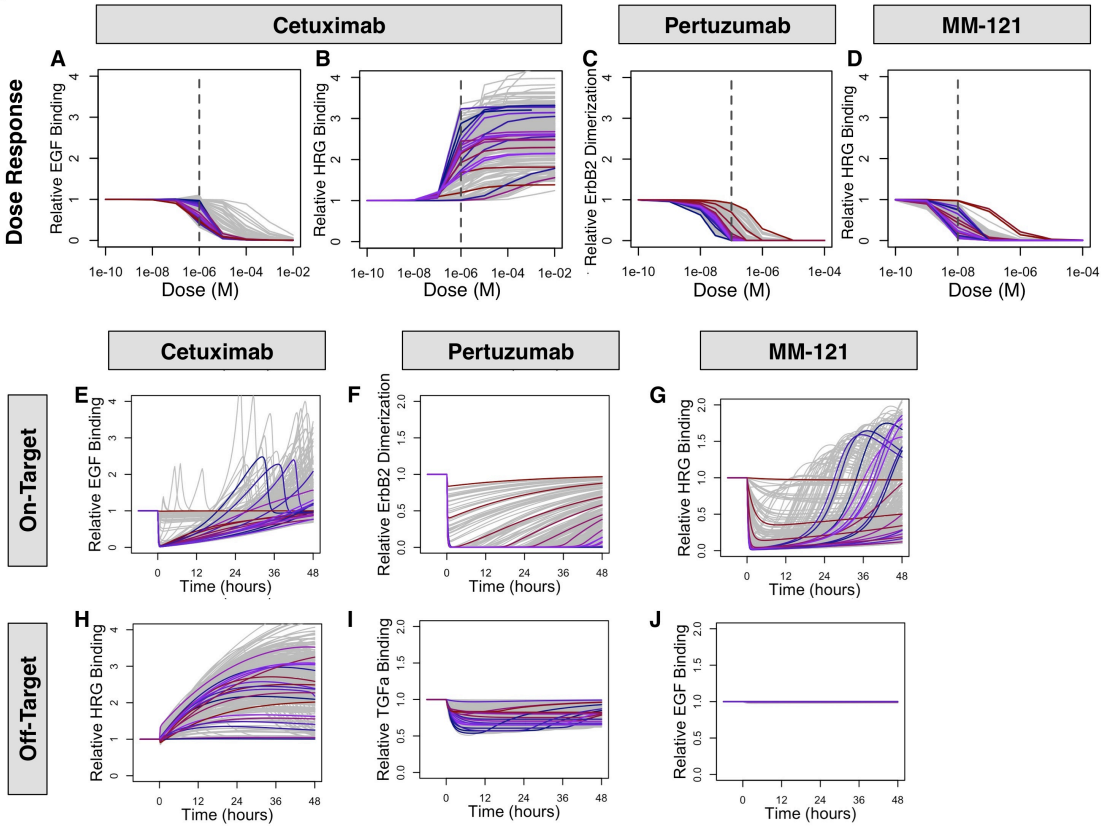

Figure 7-5: Therapeutic targeting of receptors inhibits receptor binding with some up-regulation of off-target complexes.

A-D, Dose response curves show that therapies are effective in reducing target binding – cetuximab inhibits EGF binding to EGFR (A), pertuzumab inhibits ErbB2 dimerization (C), MM-121 inhibits HRG binding to ErbB3 (D) – with some off-target effects, including the cetuximab-induced up-regulation of HRG binding to ErbB3 (B). Each line corresponds to one patient. Values on the *y*-axis correspond to the time integral of the specified complex at each dose divided by the time integral when the dose is zero. Each plot has 20 lines with non-gray colors, ranging from blue to purple to red. These samples are determined by ordering the samples according to the steady-state level of the specified complex and then choosing 20 evenly spaced samples from this list. Vertical dashed lines are located at the dose chosen for further simulations. **E-J,** Time courses at the doses selected in A-D show the varying levels of target inhibition found across the population for cetuximab (E), pertuzumab (F), and MM-121 (G), as well as off-target effects (H,

I, J). As in A-D, each line corresponds to one patient and colors are determined using the same method.

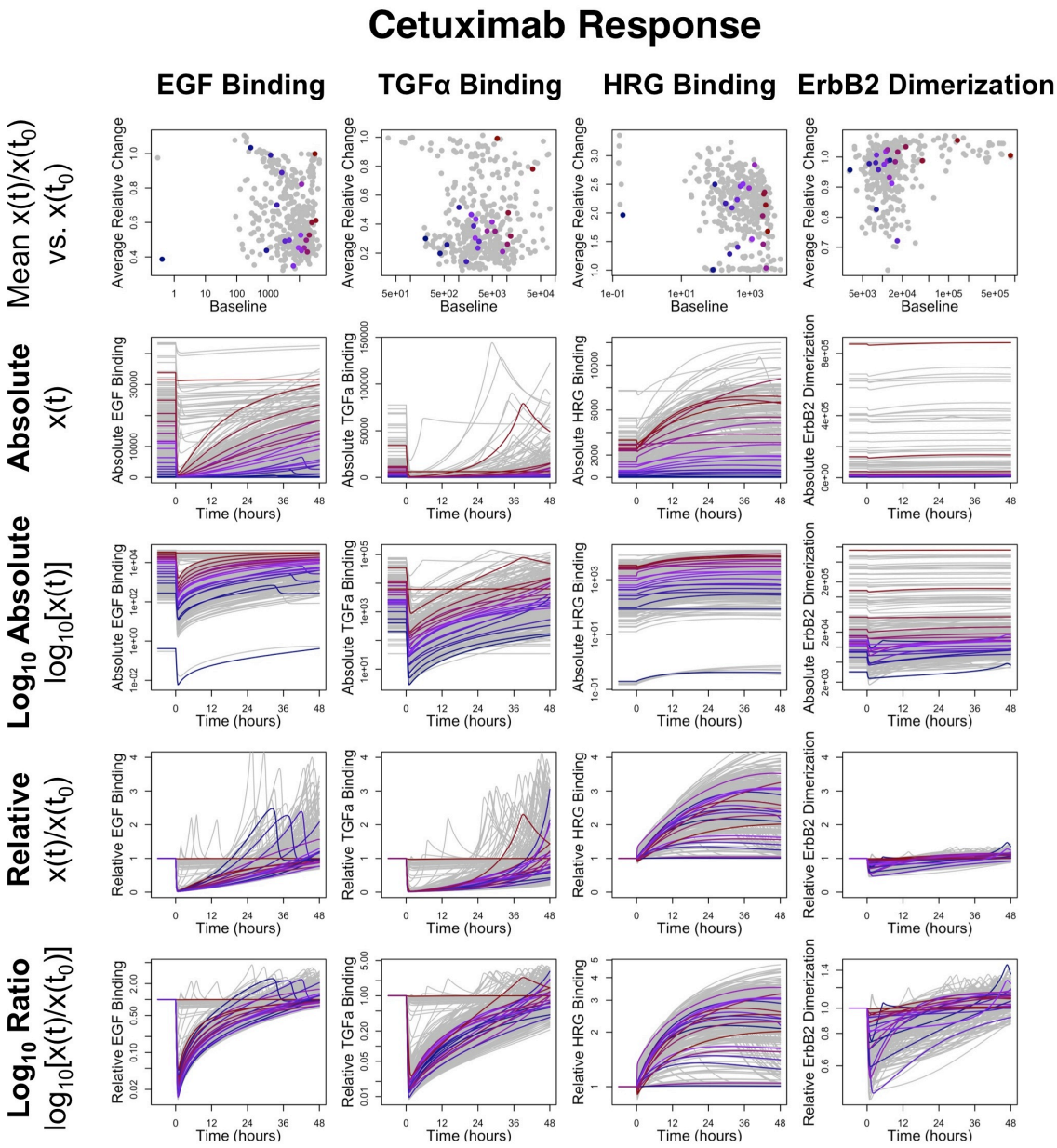

Figure 7-6: Individual cetuximab time courses.

All 400 individual time course cetuximab responses for the four reduced steady-state-derived variables. Absolute (#/cell) is shown on the top row, fractional change from baseline is shown on the bottom row.

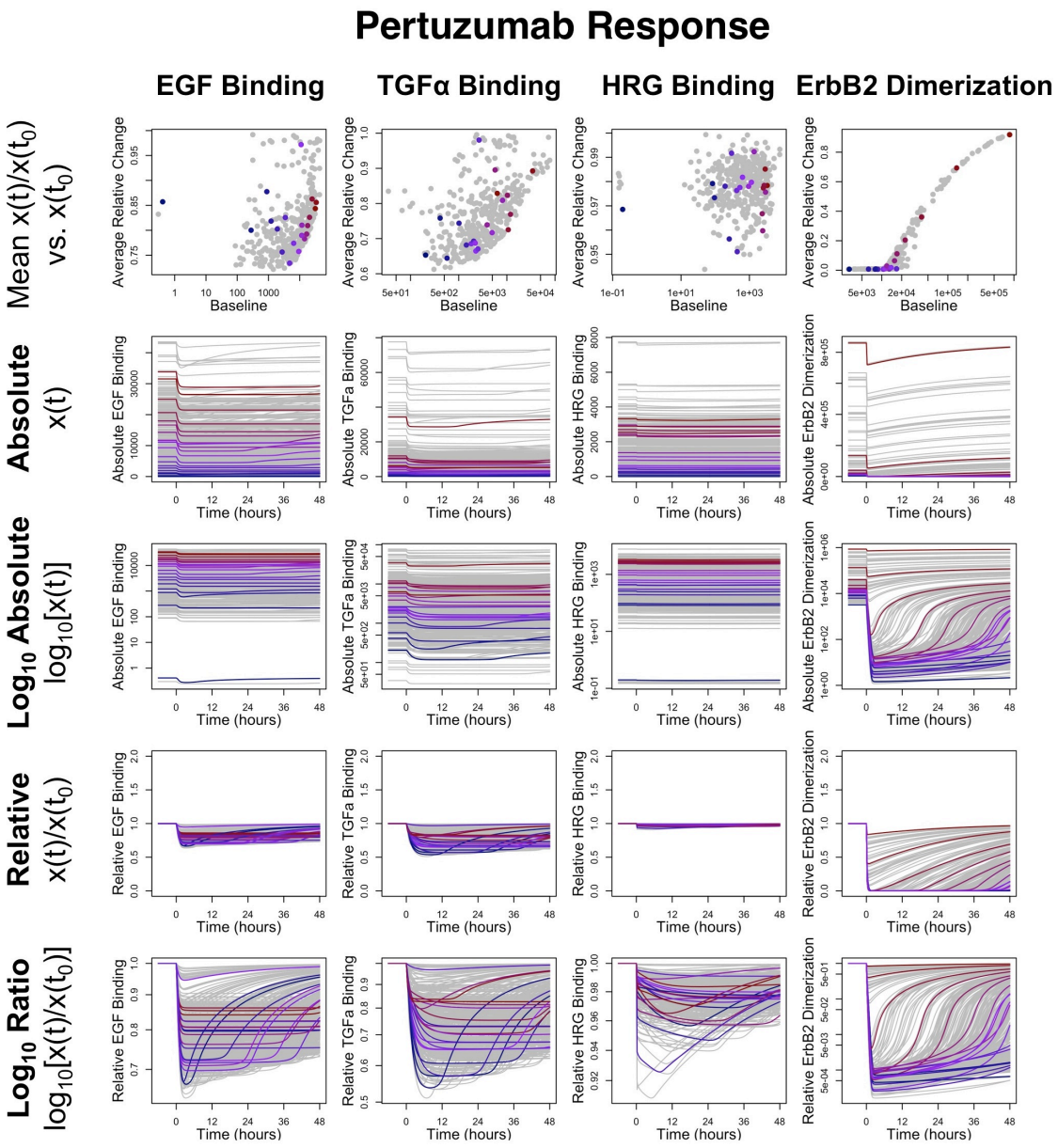

Figure 7-7: Individual pertuzumab time courses.

All 400 individual time course pertuzumab responses for the four reduced steady-state-derived variables. Absolute (#/cell) is shown on the top row, fractional change from baseline is shown on the bottom row.

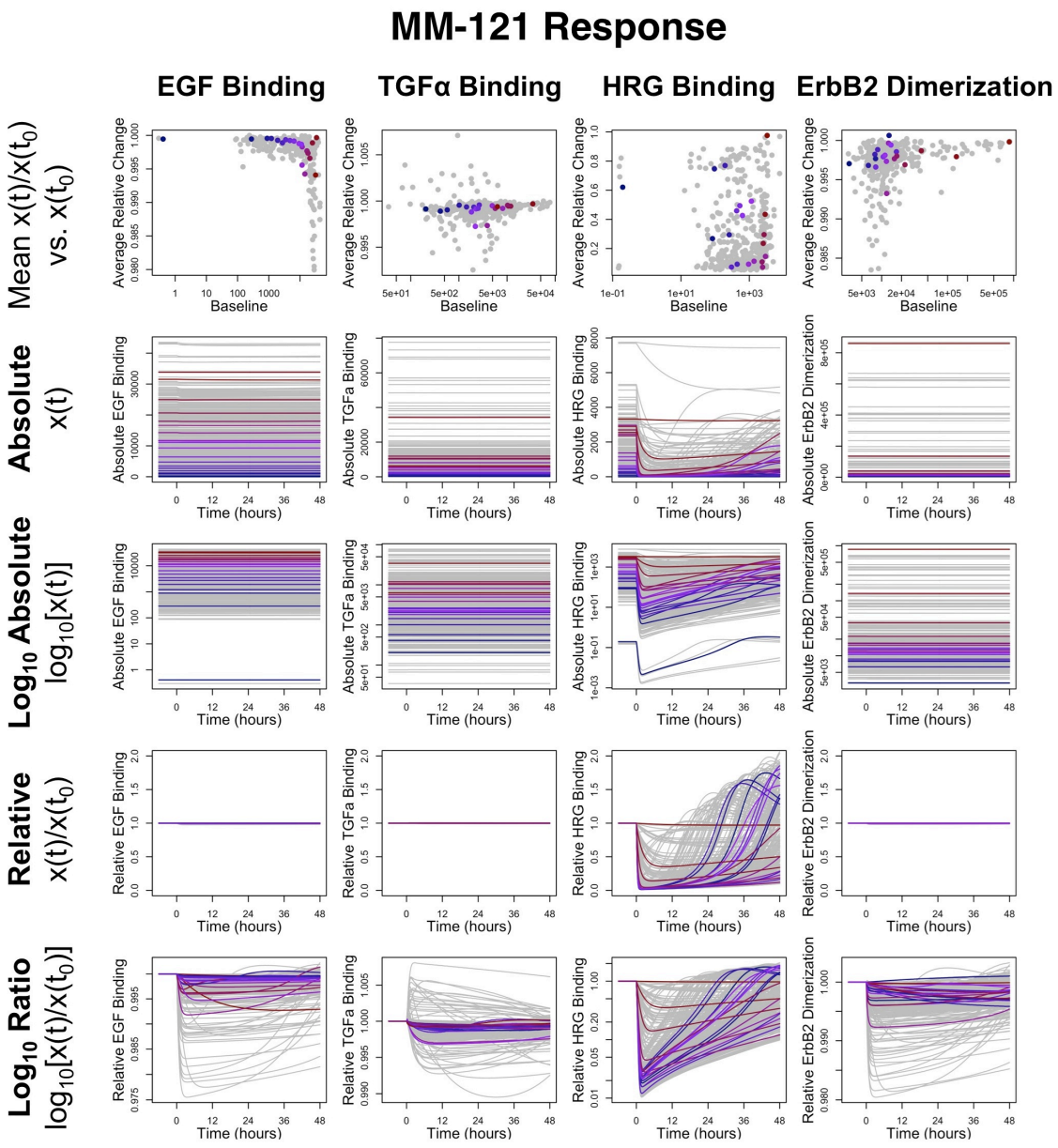

Figure 7-8: Individual MM-121 time courses.

All 400 individual time course MM-121 responses for the four reduced steady-state-derived variables. Absolute (#/cell) is shown on the top row, fractional change from baseline is shown on the bottom row.

To better assess the impact of treatments, we sought metrics that account for both on-target and off-target effects. In addition to the obvious requirement for strong on-target effects, it is also important to consider off-target effects since an unintended increase in signaling through an off-target receptor could render a therapy ineffective. To find metrics from time course drug response data in an unbiased manner, we again turned to PCA. We normalized the data by subtracting the steady-state level of each molecular complex from the time course data, and then divided by the steady-state levels. Thus the time course data was transformed so that each time point represented the fractional change from the initial steady-state level for that individual. To capture the variation in time while not including every data point (50 variables x 193 time points = 9,650 data points per sample), we averaged each variable in two time windows: 0 to 6 hours and 24 to 30 hours. This represented both the early and late responses while only requiring 100 data points per sample.

We applied PCA to each of the drug response time courses and analyzed the component loadings after varimax rotation. Although there was some evidence of gene-based clustering of variable loadings, in general the input genes did not have as strong an effect on loadings as they did in the steady-state PCA (**Figure 7-9A** for $PC1_{Cetux}$-$PC2_{Cetux}$ as an example; note the absence of color-clustering). Instead, we built aggregate variables out of collections of variables that had similar loadings and examined their time courses across the 400 patients (shown in **Figure 7-9B** for $PC1_{Cetux}$-$PC2_{Cetux}$).

For the cetuximab response, three of the principal components resulted in aggregate variables that varied significantly across the patients: $PC1_{Cetux}$ was associated with endosome EGFR complexes that were repressed for almost the entire 48-hour duration, $PC2_{Cetux}$ was associated with temporarily repressed surface EGFR complexes, and $PC4_{Cetux}$ was associated with both ErbB3 complexes and ErbB2 monomer and homodimer levels (**Figure 7-9C**). The distinction between $PC1_{Cetux}$ and $PC2_{Cetux}$ was possible because of the inclusion of both early and late time points. We used the scores for these three principal components as response metrics for

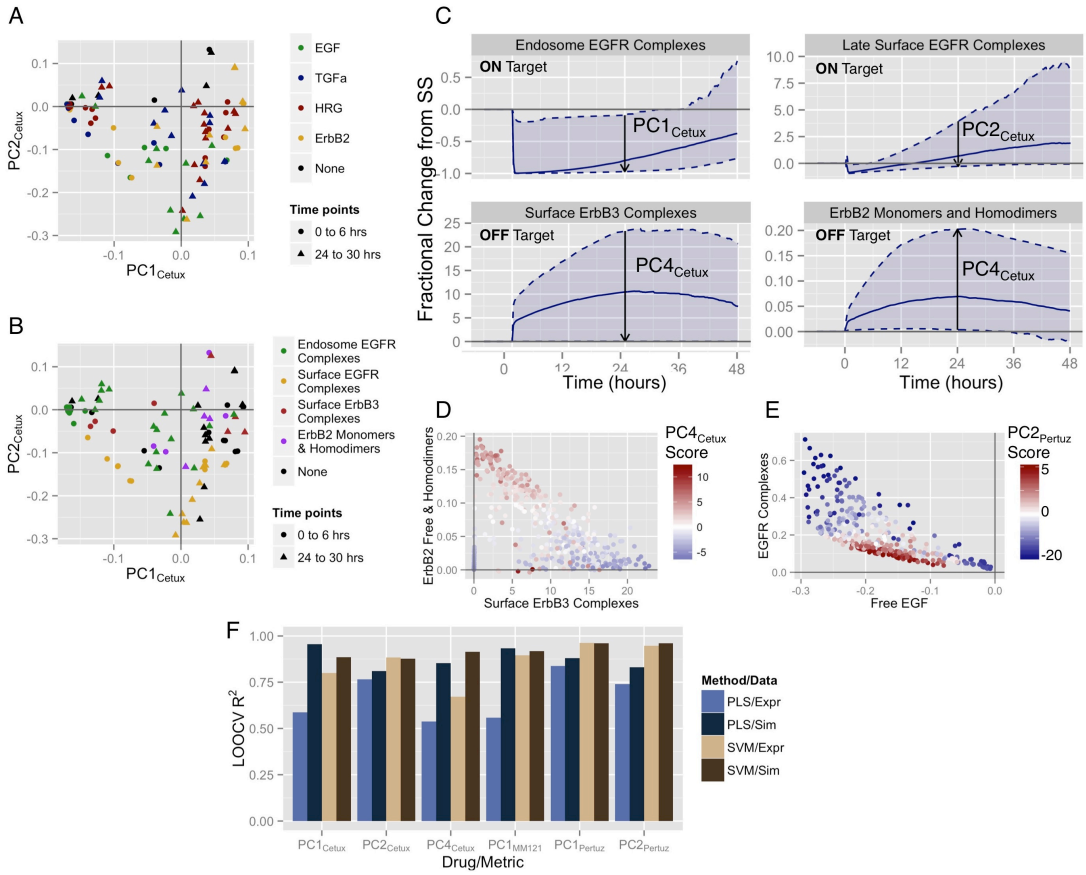

Figure 7-9: Variation in the dynamics of response to treatments is captured by PCA.

A, Projection of the PCA loadings onto the $PC1_{Cetux}$-$PC2_{Cetux}$ subspace reveals that gene expression does not linearly affect drug response time courses as it did for the steady-state simulated variables. **B**, The PCA-based aggregate variables are better able to capture the variability in cetuximab response. **C,** The four cetuximab aggregate variables with significant fractional changes from baseline include on-target reductions (endosome and late time-point surface EGFR complexes, top two panels) and off-target up-regulation (ErbB3 and ErbB2 complexes, bottom two panels). The solid line corresponds to the median across the population at each time point, while the dashed lines correspond to 5^{th} and 95^{th} percentiles. **D,** The two aggregate variables derived from $PC4_{Cetux}$ are anti-correlated. The point colors indicate the value of the $PC4_{Cetux}$ score for a given sample. **E,** The two aggregate variables derived from $PC2_{Pertuz}$ are also anti-correlated. The point colors indicate the value of the $PC2_{Pertuz}$ score for a given

sample. **F,** PLS and SVM models based on simulated variables yielded better fits than the gene expression-based models for the majority of the response metrics. Bars indicate the R^2 value obtained when comparing predictions from LOOCV to the simulated data.

cetuximab: the $PC1_{Cetux}$ and $PC2_{Cetux}$ scores captured on-target effects whereas the $PC4_{Cetux}$ score captured two off-target effects. The opposing signs of the loadings of the ErbB3 complex formation and ErbB2 monomer/dimer variables indicated that these two metrics were anti-correlated. Indeed, the $PC4_{Cetux}$ score could be divided based on these two aggregate variables (**Figure 7-9D**). For this metric, intermediate values are likely ideal, whereas extreme positive or negative values indicate drug resistance due to off-target effects.

For the pertuzumab response PCA, we found that $PC1_{Pertuz}$ and $PC2_{Pertuz}$ resulted in aggregate variables with significant variation from the baseline (data not shown). $PC1_{Pertuz}$ was associated with ErbB2 homo- and hetero-dimers, the target for this drug. The variables included in the $PC2_{Pertuz}$-based aggregate were free EGF and EGFR complexes present in the endosome, thus this represented an off-target resistance metric. The two aggregate variables with opposite loadings derived from this component displayed anti-correlation, with large increases in EGFR complexes associated with large decreases in free EGF concentration (**Figure 7-9E**). This has potential clinical utility, as the change in circulating free EGF after pertuzumab administration could be indicative of the extent of off-target up-regulation in EGFR signaling in the endosome.

For the MM121 response PCA, the only principal component that yielded aggregate variables with significant variability was $PC1_{MM121}$, which corresponded to the reduction in ErbB3 complexes (data not shown). This collection of variables was also represented in $PC4_{MM121}$, but we excluded this from further analysis due to its similarity to the first component.

7.4.4 Simulated data yields effective multivariate predictive biomarkers

We used two supervised machine learning methods, partial least squares (PLS) and support vector machines (SVM), to build regression models of the six metrics identified from the drug response PCA. We compared the performance of two types of predictor variables: gene expression data and simulated steady-state variables. To assess biomarker performance, we used a leave-one-out cross-validation (LOOCV) approach where model fitting occurred 400 times, with

one of the 400 samples withheld each time. The metric for the withheld sample was predicted, allowing us to compute the coefficient of determination (R^2) across the 400 samples.

We determined the number of components to use in the PLS regression models by finding the LOOCV R^2 for each metric at a range of component numbers. For all metrics, the PLS models with gene expression data as predictors did not improve as the number of components increased. The PLS models with simulated variables as predictors, however, did improve as the component number increased to 8 in five out of six cases (data not shown). For SVM, we used epsilon regression with a radial basis kernel with the default value for γ.

We compared the performance of the two types of predictors across both PLS and SVM models, with optimized component numbers for PLS. In almost all cases, the simulated variables outperformed the expression data as predictors (**Figure 7-9F**). In the few cases where the expression data outperformed the simulated variables, it was by only a very small margin. In general, the PLS models performed similarly to the SVM models, with very high LOOCV R^2 values especially for the three cetuximab metrics and for $PC1_{Pertuz}$.

7.4.5 Ranking of patients based on response metrics predicts similar drug response rates across breast cancer subtypes

We examined the response metrics across individuals in the four major PAM50 subtypes (basal-like/triple-negative, HER2-enriched, luminal A, and luminal B) to determine whether any of these subtypes were more susceptible to the drugs being tested. Ordering the samples within each subtype by the on-target metric produced no patterns in the off-target metrics, indicating a lack of correlation between the on-target and off-target responses. There were also no subtype-specific patterns in either the on- or off-target responses (**Figure 7-10**); each of the subtypes were comprised of both likely responders and likely non-responders. This suggests that identification

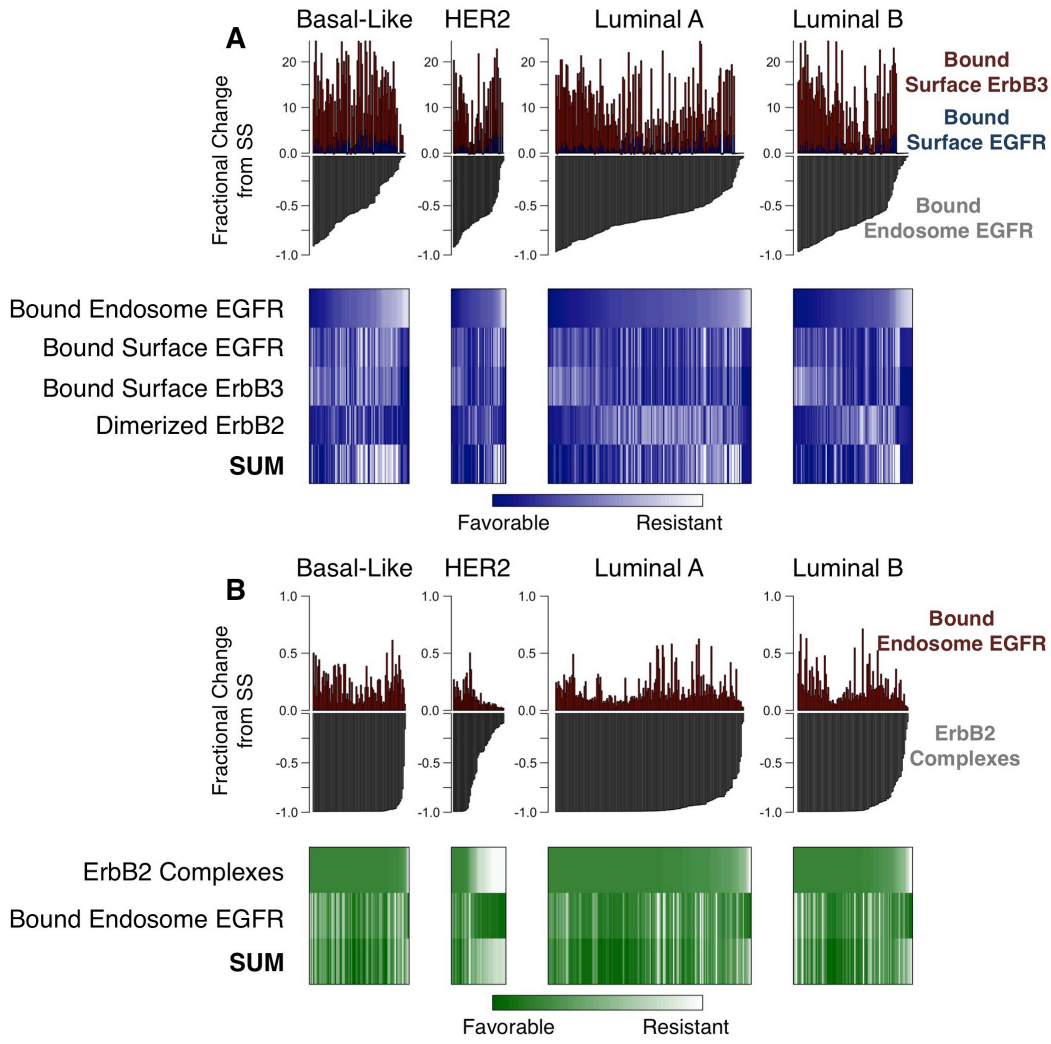

Figure 7-10: Treatment response metrics further subdivide PAM50 subtypes into treatment sensitive and resistant groups.

A, Average fractional changes from steady-state (top panel) in each individual of endosome EGFR binding (negative axis) and the three cetuximab resistance mechanisms, namely up-regulation of surface EGFR binding (blue), surface ErbB3 binding (red), and ErbB2 dimerization (does not appear due to low magnitude). Patients are split into the four major PAM50 subtypes and are ordered by the magnitude of down-regulation of endosome EGFR binding. The heatmap (bottom panel) shows scaled fractional changes in the cetuximab response data. Individuals are ordered as in the top panel, but now the scaling allows each of the four metrics to be represented

on the same color scale. Darker blue indicates a more favorable outcome, either more reduction in endosome EGFR binding or smaller increases in the other three metrics. **B,** As for A, but with the pertuzumab response data. Individuals are now ordered by the magnitude of reduction in ErbB2 dimerization. The other metric included is endosome EGFR complexes (red; always increased in response to pertuzumab).

of subtype, which is itself highly correlated with gene expression clustering and pre-treatment signaling levels (**Figure 7-4C**), is insufficient for predicting response to drugs. The nonlinear nature of the protein interaction networks being targeted by these drugs results in nonlinear effects of the drugs that run across the subtypes. Incorporation of mechanistic, quantitative interaction information gives hope that better predictions of drug effects can be made.

7.5 Discussion

Here we showed that combining tumor measurements of EGF family gene expression with computational kinetic models (based on quantitative mechanistic knowledge) can enhance the prediction of drug response. In addition to predictive biomarkers for EGF family-targeting drugs, our approach also yields potential insight into mechanisms of resistance to these treatments and shows that the traditional basal/luminal distinction of breast tumors may not be sufficient in making treatment decisions regarding EGF family-targeting drugs. One limitation of this approach is that gene and protein expression measurements are expected to be noisy and may not always correlate with actual *in vivo* protein production rates. Another limitation is that the use of pre-treatment gene expression in the model can account for intrinsic resistance to therapies but not adaptive resistance, which includes longer-term changes arising due to adjustment of protein production rates or activation of entirely different pathways. Nonetheless, the approach described here could yield biomarkers with clinical utility in treatment selection, and can be adapted to obtain a fuller picture of the therapeutic dynamics, by incorporating post-treatment measurements of expression over time.

Complex biomarkers involving as many as 100 genes have been developed, however these linear combinations do not take advantage of the hard-won molecular-mechanism information available for the drug targets and their interacting proteins. By using our knowledge of molecular mechanisms of interaction, as well as the individualized gene and protein expression levels of multiple ligands and receptors, we can predict mechanism-based nonlinear latent

variables that could not be obtained from expression data alone, and which may be more effective biomarkers.

We found evidence in literature for several of the mechanisms of resistance described in this study. In colorectal cancer patients treated with cetuximab, both of the resistance mechanisms associated with the $PC4_{Cetux}$ metric have been observed: patients with ErbB2 amplification or high plasma levels of HRG have lower overall survival after cetuximab treatment [170]. We were unable to find clinical data implicating EGFR and ErbB3 ligands in pertuzumab resistance, but one study found that trastuzumab-resistant breast cancer xenografts expressed higher levels of TGFα, HBEGF, and HRG than sensitive xenografts [181].

Our finding that transcriptomic and proteomic data had strong linear relationships with baseline pre-treatment EGF family signaling, but weaker relationships with post-treatment response, suggests a possible reason why gene expression-based subtypes often have overall prognostic significance but fail to predict treatment response. Based on our analysis, gene expression measurements can describe the baseline signaling activity; however, when a treatment is added, the complex molecular interactions make it difficult to predict the response of the pathway without the help of a computational model. This discrepancy in the predictive capability of baseline and stimulated signaling has been suggested previously: one study found that data describing the response of breast cancer cells to HRG provided better prediction of drug response to phosphoinositide 3-kinase (PI3K) inhibitors than baseline receptor phosphorylation levels [182].

The approach described here to explore drug response metrics can be extended to other drugs and growth factor families, and could also include simulating alternate dose scheduling, routes of administration and drug release profiles. To validate the approach, paired genomic and clinical outcome data is needed; the samples included in this study did not have information regarding clinical outcomes. An ideal dataset would consist of gene expression data from biopsy samples paired with the clinical response to a drug (e.g. complete or partial response versus stable

disease or progression), and would be further enhanced by the obtaining of multiple samples over time to improve simulation of the response.

8 Conclusions and Future Directions

8.1 Mechanism-Based Gene Selection for Biomarkers

We used multivariate statistical methods to identify groups of patients with altered expression of VEGF- and Semaphorin-related genes. In breast cancer, a signature consisting of high *VEGFA* expression and low class 3 Semaphorin expression was associated with the aggressive triple-negative subtype, suggesting a role for angiogenesis due to elevated VEGF and reduced anti-angiogenic Semaphorins. Analysis of the survival times in different VEGF/Sema clusters revealed that the high *VEGFA*/low class 3 Sema pattern, although associated with TNBCs, also had significant prognostic effects in non-TNBCs. Thus, it identified a group of patients who were not triple-negative but had poor outcomes similar to triple-negative patients. Similarly, within the HER2-enriched PAM50 subtype, the high VEGFA/low class 3 Sema pattern was associated with poor prognosis, whereas HER2-enriched patients not enriched for this pattern had good prognoses. Clustering the VEGF/Sema gene expression of 2,656 breast cancer samples revealed strong overlap between the VEGF/Sema clusters and the PAM50 subtypes. Clustering only the triple-negative breast cancers resulted in five distinct clusters. One of these clusters was associated with one of the six previously-defined TNBC subtypes [81], the mesenchymal stem-like subtype, which is enriched with expression of epithelial-to-mesenchymal transition genes. The pattern of VEGF-related gene expression found in this subtype was elevated *VEGFC*, *KDR*, and *NRP1*.

TCGA gene expression datasets for other cancer types did not display the high *VEGFA*/low class 3 Semaphorin expression pattern, but the elevated *VEGFC*, *KDR*, and *NRP1* pattern was found in several subtypes of cancers, notably the MSI/CIMP subtype of colon cancer and the mesenchymal subtype of glioblastoma. The VEGF/Sema cluster with this pattern in glioblastoma had significantly worse survival times than other clusters, suggesting a role in disease

progression. Notably, out of the seven TCGA primary tumor datasets analyzed, only prostate cancer failed to yield distinct VEGF/Sema clusters. Thus, we explored prostate cancer further.

In several prostate cancer microarray and RNA-Seq datasets, we found that both *VEGFA* and class 3 semaphorins were down-regulated in primary prostate tumors compared to normal prostate tissue. The reduced class 3 Semaphorin expression suggests removal of inhibitory signals for angiogenesis, while the reduced *VEGFA* expression suggests a lack of responsiveness to anti-VEGF therapies. We then examined gene expression datasets for *metastatic* prostate cancer and found that *VEGFA* was up-regulated, along with down-regulation of class 3 semaphorins. Thus, although VEGF-based subtypes of primary prostate cancers are difficult to detect, metastatic prostate cancer may be responsive to VEGF-targeting therapies. Simulations of VEGF signaling with protein production rates proportional to gene expression will provide further information regarding the role of elevated *VEGFA* and reduced class 3 Semaphorins.

8.2 Personalized Computational Models for Biomarker Testing

We used computational modeling to predict growth factor signaling and drug response for genes that encode proteins whose interactions and kinetics were known. We applied this in multiple settings.

First, we considered a single compartment model where the compartment corresponded to normal prostate tissue, a primary prostate tumor, or a metastasis originating from a prostate tumor. This model showed that VEGF-Semaphorin competition for Neuropilin co-receptors likely has little impact on either VEGF signaling or Semaphorin signaling. Thus, class 3 Semaphorin-mediated anti-angiogenic activity likely occurs due to signals transduced by plexin receptors after interacting with Sema3-NRP complexes. We found that some metastases had down-regulated class 3 Semaphorin signaling, others had up-regulated VEGF signaling, and others had both. This demonstrates the need for considering multiple pathways when targeting angiogenesis.

We then analyzed VEGF signaling in whole-body multi-compartment models based on gene expression from three TCGA datasets: breast, kidney, and prostate cancers. We found that expression of genes encoding soluble ligands were highly correlated with the untreated baseline level of the corresponding ligand-receptor complexes. This was not the case when comparing gene expression to the changes in ligand-receptor complexes after treatment with either bevacizumab (anti-VEGF) or anti-NRP1. When considering gene expression variables as predictors of treatment response, we found that different genes served as the best predictors in different types of cancer. This demonstrates that differing covariances between genes in different cancer types can lead to changes in how predictive of response a particular biomarker is. This is seen in clinical trials of bevacizumab, where plasma VEGFA concentration is only predictive of response for some cancers [39]. The cell type that exhibits the most variability in receptor levels (tumor versus endothelial) can also affect biomarker performance. Across the virtual populations, simulated baseline variables tended to outperform gene expression variables as predictive biomarkers, suggesting that the incorporation of mechanistic information does yield improved biomarkers.

We also applied our method to the well-studied EGFR/ErbB family of growth factor receptors. Using a previously published model of ligand-receptor interactions, with some key additions, we tested the effects of cetuximab (anti-EGFR), pertuzumab (anti-ErbB2), and MM-121 (anti-ErbB3) in a virtual population derived from the TCGA breast cancer RNA-Seq dataset. We predicted that cetuximab would have off-target effects (increased HRG-ErbB3 binding and in some cases, increased EGFR binding at the cell surface) that could lead to treatment resistance. We also predicted that pertuzumab may lead to elevated EGFR signaling in the endosome, but we found no evidence of MM-121 off-target effects. Biomarkers for treatment response metrics typically had better performance when they were built with simulated baseline data instead of gene expression data. This was consistent with our results for the VEGF signaling compartment

models. Together, this suggests that biomarkers from computational models can improve drug response prediction.

8.3 Future Directions

One potential future direction for this work is to incorporate additional data. Measurements of cytokine protein levels, either in the plasma or within tumors, would be useful in further constraining the model to ensure that the parameter values are accurate and yield robust performance. In the gene expression-based virtual populations, clinical outcome data in response to the drugs of interest would be helpful. Of particular interest would be post-treatment clinical outcome data paired with a transcriptomic dataset taken from the tumor. This would allow us to assess whether the on-target and off-target drug response metrics that we derived are truly associated with patient responses.

The framework we have developed also lends itself to application to other growth factor families with roles in angiogenesis. Many protein families other than VEGF and Semaphorin contribute to angiogenesis. We considered the platelet-derived growth factor (PDGF), insulin-like growth factor (IGF), hepatocyte growth factor (HGF), epidermal growth factor (EGF), and angiopoietin families in Chapter 4. These and other families may be able to support angiogenesis independently of the VEGF family. Ultimately, once more measurements and mechanistic information become available for these pathways, large comprehensive models such as the one developed here may enable a comprehensive analysis of the multiple families that affect tumor angiogenesis.

The compartment model can also be extended to include additional compartments corresponding to a primary tumor and multiple metastases, which is often the setting in which anti-VEGF therapeutics are administered. Measurements describing the geometric characteristics and permeability of the metastases would be needed. This would make the models more realistic and allow for improved prediction of drug response.

References

1. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144: 646-674.

2. Welti J, Loges S, Dimmeler S, Carmeliet P (2013) Recent molecular discoveries in angiogenesis and antiangiogenic therapies in cancer. J Clin Invest 123: 3190-3200.

3. Watnick RS (2012) The role of the tumor microenvironment in regulating angiogenesis. Cold Spring Harb Perspect Med 2: a006676.

4. Hanahan D, Folkman J (1996) Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. Cell 86: 353-364.

5. Lawler PR, Lawler J (2012) Molecular basis for the regulation of angiogenesis by thrombospondin-1 and -2. Cold Spring Harb Perspect Med 2: a006627.

6. Koskimaki JE, Karagiannis ED, Rosca EV, Vesuna F, Winnard PT, Jr., et al. (2009) Peptides derived from type IV collagen, CXC chemokines, and thrombospondin-1 domain-containing proteins inhibit neovascularization and suppress tumor growth in MDA-MB-231 breast cancer xenografts. Neoplasia 11: 1285-1291.

7. Koch S, Claesson-Welsh L (2012) Signal transduction by vascular endothelial growth factor receptors. Cold Spring Harb Perspect Med 2: a006502.

8. Koch S, Tugues S, Li X, Gualandi L, Claesson-Welsh L (2011) Signal transduction by vascular endothelial growth factor receptors. Biochem J 437: 169-183.

9. Neufeld G, Sabag AD, Rabinovicz N, Kessler O (2012) Semaphorins in angiogenesis and tumor progression. Cold Spring Harb Perspect Med 2: a006718.

10. Maione F, Molla F, Meda C, Latini R, Zentilin L, et al. (2009) Semaphorin 3A is an endogenous angiogenesis inhibitor that blocks tumor growth and normalizes tumor vasculature in transgenic mouse models. J Clin Invest 119: 3356-3372.

11. Varshavsky A, Kessler O, Abramovitch S, Kigel B, Zaffryar S, et al. (2008) Semaphorin-3B is an angiogenesis inhibitor that is inactivated by furin-like pro-protein convertases. Cancer Res 68: 6922-6931.

12. Kigel B, Varshavsky A, Kessler O, Neufeld G (2008) Successful inhibition of tumor development by specific class-3 semaphorins is associated with expression of appropriate semaphorin receptors by tumor cells. PLoS One 3: e3287.

13. Toyofuku T, Yabuki M, Kamei J, Kamei M, Makino N, et al. (2007) Semaphorin-4A, an activator for T-cell-mediated immunity, suppresses angiogenesis via Plexin-D1. EMBO J 26: 1373-1384.

14. Basile JR, Afkhami T, Gutkind JS (2005) Semaphorin 4D/plexin-B1 induces endothelial cell migration through the activation of PYK2, Src, and the phosphatidylinositol 3-kinase-Akt pathway. Mol Cell Biol 25: 6889-6898.

15. Sadanandam A, Rosenbaugh EG, Singh S, Varney M, Singh RK (2010) Semaphorin 5A promotes angiogenesis by increasing endothelial cell proliferation, migration, and decreasing apoptosis. Microvasc Res 79: 1-9.

16. Miao HQ, Soker S, Feiner L, Alonso JL, Raper JA, et al. (1999) Neuropilin-1 mediates collapsin-1/semaphorin III inhibition of endothelial cell motility: functional competition of collapsin-1 and vascular endothelial growth factor-165. J Cell Biol 146: 233-242.

17. Gu C, Limberg BJ, Whitaker GB, Perman B, Leahy DJ, et al. (2002) Characterization of neuropilin-1 structural features that confer binding to semaphorin 3A and vascular endothelial growth factor 165. J Biol Chem 277: 18069-18076.

18. Guo HF, Li X, Parker MW, Waltenberger J, Becker PM, et al. (2013) Mechanistic basis for the potent anti-angiogenic activity of semaphorin 3F. Biochemistry 52: 7551-7558.

19. Parker MW, Linkugel AD, Vander Kooi CW (2013) Effect of C-terminal sequence on competitive semaphorin binding to neuropilin-1. J Mol Biol 425: 4405-4414.

20. Hurwitz H, Fehrenbacher L, Novotny W, Cartwright T, Hainsworth J, et al. (2004) Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. N Engl J Med 350: 2335-2342.

21. Sandler A, Gray R, Perry MC, Brahmer J, Schiller JH, et al. (2006) Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. N Engl J Med 355: 2542-2550.

22. Vredenburgh JJ, Desjardins A, Reardon DA, Peters KB, Herndon JE, 2nd, et al. (2011) The addition of bevacizumab to standard radiation therapy and temozolomide followed by bevacizumab, temozolomide, and irinotecan for newly diagnosed glioblastoma. Clin Cancer Res 17: 4119-4124.

23. Friedman HS, Prados MD, Wen PY, Mikkelsen T, Schiff D, et al. (2009) Bevacizumab alone and in combination with irinotecan in recurrent glioblastoma. J Clin Oncol 27: 4733-4740.

24. Escudier B, Pluzanska A, Koralewski P, Ravaud A, Bracarda S, et al. (2007) Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. Lancet 370: 2103-2111.

25. Escudier B, Bellmunt J, Negrier S, Bajetta E, Melichar B, et al. (2010) Phase III trial of bevacizumab plus interferon alfa-2a in patients with metastatic renal cell carcinoma (AVOREN): final analysis of overall survival. J Clin Oncol 28: 2144-2150.

26. Tewari KS, Sill MW, Long HJ, 3rd, Penson RT, Huang H, et al. (2014) Improved survival with bevacizumab in advanced cervical cancer. N Engl J Med 370: 734-743.

27. Kindler HL, Niedzwiecki D, Hollis D, Sutherland S, Schrag D, et al. (2010) Gemcitabine plus bevacizumab compared with gemcitabine plus placebo in patients with advanced pancreatic cancer: phase III trial of the Cancer and Leukemia Group B (CALGB 80303). J Clin Oncol 28: 3617-3622.

28. Miller KD, Chap LI, Holmes FA, Cobleigh MA, Marcom PK, et al. (2005) Randomized phase III trial of capecitabine compared with bevacizumab plus capecitabine in patients with previously treated metastatic breast cancer. J Clin Oncol 23: 792-799.

29. Miller K, Wang M, Gralow J, Dickler M, Cobleigh M, et al. (2007) Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. N Engl J Med 357: 2666-2676.

30. Perren TJ, Swart AM, Pfisterer J, Ledermann JA, Pujade-Lauraine E, et al. (2011) A phase 3 trial of bevacizumab in ovarian cancer. N Engl J Med 365: 2484-2496.

31. Kelly WK, Halabi S, Carducci M, George D, Mahoney JF, et al. (2012) Randomized, double-blind, placebo-controlled phase III trial comparing docetaxel and prednisone with or without bevacizumab in men with metastatic castration-resistant prostate cancer: CALGB 90401. J Clin Oncol 30: 1534-1540.

32. Fuchs CS, Tomasek J, Yong CJ, Dumitru F, Passalacqua R, et al. (2014) Ramucirumab monotherapy for previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (REGARD): an international, randomised, multicentre, placebo-controlled, phase 3 trial. Lancet 383: 31-39.

33. Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, et al. (2007) Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. N Engl J Med 356: 115-124.

34. Raymond E, Dahan L, Raoul JL, Bang YJ, Borbath I, et al. (2011) Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. N Engl J Med 364: 501-513.

35. Demetri GD, van Oosterom AT, Garrett CR, Blackstein ME, Shah MH, et al. (2006) Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomised controlled trial. Lancet 368: 1329-1338.

36. Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, et al. (2008) Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 359: 378-390.

37. Ohtsu A, Shah MA, Van Cutsem E, Rha SY, Sawaki A, et al. (2011) Bevacizumab in combination with chemotherapy as first-line therapy in advanced gastric cancer: a randomized, double-blind, placebo-controlled phase III study. J Clin Oncol 29: 3968-3976.

38. Cameron D, Brown J, Dent R, Jackisch C, Mackey J, et al. (2013) Adjuvant bevacizumab-containing therapy in triple-negative breast cancer (BEATRICE): primary results of a randomised, phase 3 trial. Lancet Oncol 14: 933-942.

39. Lambrechts D, Lenz HJ, de Haas S, Carmeliet P, Scherer SJ (2013) Markers of response for the antiangiogenic agent bevacizumab. J Clin Oncol 31: 1219-1230.

40. Miles DW, de Haas SL, Dirix LY, Romieu G, Chan A, et al. (2013) Biomarker results from the AVADO phase 3 trial of first-line bevacizumab plus docetaxel for HER2-negative metastatic breast cancer. Br J Cancer 108: 1052-1060.

41. Van Cutsem E, de Haas S, Kang YK, Ohtsu A, Tebbutt NC, et al. (2012) Bevacizumab in combination with chemotherapy as first-line therapy in advanced gastric cancer: a biomarker evaluation from the AVAGAST randomized phase III trial. J Clin Oncol 30: 2119-2127.

42. Mac Gabhann F, Popel AS (2004) Model of competitive binding of vascular endothelial growth factor and placental growth factor to VEGF receptors on endothelial cells. Am J Physiol Heart Circ Physiol 286: H153-164.

43. Mac Gabhann F, Popel AS (2005) Differential binding of VEGF isoforms to VEGF receptor 2 in the presence of neuropilin-1: a computational model. Am J Physiol Heart Circ Physiol 288: H2851-2860.

44. Stefanini MO, Wu FT, Mac Gabhann F, Popel AS (2008) A compartment model of VEGF distribution in blood, healthy and diseased tissues. BMC Syst Biol 2: 77.

45. Wu FT, Stefanini MO, Mac Gabhann F, Kontos CD, Annex BH, et al. (2009) Computational kinetic model of VEGF trapping by soluble VEGF receptor-1: effects of transendothelial and lymphatic macromolecular transport. Physiol Genomics 38: 29-41.

46. Stefanini MO, Wu FT, Mac Gabhann F, Popel AS (2010) Increase of plasma VEGF after intravenous administration of bevacizumab is predicted by a pharmacokinetic model. Cancer Res 70: 9886-9894.

47. Finley SD, Engel-Stefanini MO, Imoukhuede PI, Popel AS (2011) Pharmacokinetics and pharmacodynamics of VEGF-neutralizing antibodies. BMC Syst Biol 5: 193.

48. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98: 10869-10874.

49. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. (2000) Molecular portraits of human breast tumours. Nature 406: 747-752.

50. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, et al. (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 27: 1160-1167.

51. (2012) Comprehensive molecular portraits of human breast tumours. Nature 490: 61-70.

52. Foulkes WD, Smith IE, Reis-Filho JS (2010) Triple-negative breast cancer. N Engl J Med 363: 1938-1948.

53. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, et al. (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 17: 98-110.

54. (2012) Comprehensive molecular characterization of human colon and rectal cancer. Nature 487: 330-337.

55. (2011) Integrated genomic analyses of ovarian carcinoma. Nature 474: 609-615.

56. Tan-Chiu E, Yothers G, Romond E, Geyer CE, Jr., Ewer M, et al. (2005) Assessment of cardiac dysfunction in a randomized trial comparing doxorubicin and cyclophosphamide

followed by paclitaxel, with or without trastuzumab as adjuvant therapy in node-positive, human epidermal growth factor receptor 2-overexpressing breast cancer: NSABP B-31. J Clin Oncol 23: 7811-7819.

57. Baselga J, Cortes J, Kim SB, Im SA, Hegg R, et al. (2012) Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. N Engl J Med 366: 109-119.

58. Cameron D, Casey M, Oliva C, Newstat B, Imwalle B, et al. (2010) Lapatinib plus capecitabine in women with HER-2-positive advanced breast cancer: final survival analysis of a phase III randomized trial. Oncologist 15: 924-934.

59. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, et al. (2011) Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med 364: 2507-2516.

60. Paik S, Tang G, Shak S, Kim C, Baker J, et al. (2006) Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol 24: 3726-3734.

61. Pawitan Y, Bjohle J, Amler L, Borg AL, Egyhazi S, et al. (2005) Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. Breast Cancer Res 7: R953-964.

62. Farmer P, Bonnefoi H, Becette V, Tubiana-Hulin M, Fumoleau P, et al. (2005) Identification of molecular apocrine breast tumours by microarray analysis. Oncogene 24: 4660-4671.

63. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, et al. (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 365: 671-679.

64. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, et al. (2005) Genes that mediate breast cancer metastasis to lung. Nature 436: 518-524.

65. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, et al. (2006) Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 98: 262-272.

66. Miller LD, Smeds J, George J, Vega VB, Vergara L, et al. (2005) An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci U S A 102: 13550-13555.

67. Minn AJ, Gupta GP, Padua D, Bos P, Nguyen DX, et al. (2007) Lung metastasis genes couple breast tumor size and metastatic spread. Proc Natl Acad Sci U S A 104: 6740-6745.

68. Boersma BJ, Reimers M, Yi M, Ludwig JA, Luke BT, et al. (2008) A stromal gene signature associated with inflammatory breast cancer. Int J Cancer 122: 1324-1332.

69. Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, et al. (2007) Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res 13: 3207-3214.

70. Schmidt M, Bohm D, von Torne C, Steiner E, Puhl A, et al. (2008) The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer Res 68: 5405-5413.

71. Popovici V, Chen W, Gallas BG, Hatzis C, Shi W, et al. (2010) Effect of training-sample size and classification difficulty on the accuracy of genomic predictors. Breast Cancer Res 12: R5.

72. Tabchy A, Valero V, Vidaurre T, Lluch A, Gomez H, et al. (2010) Evaluation of a 30-gene paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide chemotherapy response predictor in a multicenter randomized trial in breast cancer. Clin Cancer Res 16: 5351-5361.

73. Graham K, de las Morenas A, Tripathi A, King C, Kavanah M, et al. (2010) Gene expression in histologically normal epithelium from breast cancer patients and from cancer-free prophylactic mastectomy patients shares a similar profile. Br J Cancer 102: 1284-1293.

74. Palaskas N, Larson SM, Schultz N, Komisopoulou E, Wong J, et al. (2011) 18F-fluorodeoxy-glucose positron emission tomography marks MYC-overexpressing human basal-like breast cancers. Cancer Res 71: 5164-5174.

75. Iwamoto T, Bianchini G, Booser D, Qi Y, Coutant C, et al. (2011) Gene pathways associated with prognosis and chemotherapy sensitivity in molecular subtypes of breast cancer. J Natl Cancer Inst 103: 264-272.

76. Iwamoto T, Bianchini G, Qi Y, Cristofanilli M, Lucci A, et al. (2011) Different gene expressions are associated with the different molecular subtypes of inflammatory breast cancer. Breast Cancer Res Treat 125: 785-795.

77. Creighton CJ, Sada YH, Zhang Y, Tsimelzon A, Wong H, et al. (2012) A gene transcription signature of obesity in breast cancer. Breast Cancer Res Treat 132: 993-1000.

78. Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, et al. (2011) A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. JAMA 305: 1873-1881.

79. Karn T, Pusztai L, Holtrich U, Iwamoto T, Shiang CY, et al. (2011) Homogeneous datasets of triple negative breast cancers enable the identification of novel prognostic and predictive signatures. PLoS One 6: e28403.

80. Gonzalez-Angulo AM, Iwamoto T, Liu S, Chen H, Do KA, et al. (2012) Gene expression, molecular class changes, and pathway analysis after neoadjuvant systemic therapy for breast cancer. Clin Cancer Res 18: 1109-1119.

81. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, et al. (2011) Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. Journal of Clinical Investigation 121: 2750-2767.

82. Monti S, Tamayo P, Mesirov J, Golub T (2003) Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. Machine Learning 52: 91-118.

83. Guttmann-Raviv N, Shraga-Heled N, Varshavsky A, Guimaraes-Sternberg C, Kessler O, et al. (2007) Semaphorin-3A and Semaphorin-3F Work Together to Repel Endothelial Cells and to Inhibit Their Survival by Induction of Apoptosis. Journal of Biological Chemistry 282: 26294-26305.

84. Esselens C, Malapeira J, Colome N, Casal C, Rodriguez-Manzaneque JC, et al. (2010) The cleavage of semaphorin 3C induced by ADAMTS1 promotes cell migration. J Biol Chem 285: 2463-2473.

85. Banu N, Teichman J, Dunlap-Brown M, Villegas G, Tufro A (2006) Semaphorin 3C regulates endothelial cell function by increasing integrin activity. FASEB J 20: 2150-2152.

86. Gu C, Yoshida Y, Livet J, Reimert DV, Mann F, et al. (2005) Semaphorin 3E and plexin-D1 control vascular pattern independently of neuropilins. Science 307: 265-268.

87. Sakurai A, Gavard J, Annas-Linhares Y, Basile JR, Amornphimoltham P, et al. (2010) Semaphorin 3E initiates antiangiogenic signaling through plexin D1 by regulating Arf6 and R-Ras. Mol Cell Biol 30: 3086-3098.

88. Kessler O, Shraga-Heled N, Lange T, Gutmann-Raviv N, Sabo E, et al. (2004) Semaphorin-3F is an inhibitor of tumor angiogenesis. Cancer Res 64: 1008-1015.

89. Parker MW, Hellman LM, Xu P, Fried MG, Vander Kooi CW (2010) Furin processing of semaphorin 3F determines its anti-angiogenic activity by regulating direct binding and competition for neuropilin. Biochemistry 49: 4068-4075.

90. Kutschera S, Weber H, Weick A, De Smet F, Genove G, et al. (2011) Differential endothelial transcriptomics identifies semaphorin 3G as a vascular class 3 semaphorin. Arterioscler Thromb Vasc Biol 31: 151-159.

91. Basile JR (2006) Semaphorin 4D provides a link between axon guidance processes and tumor-induced angiogenesis. Proceedings of the National Academy of Sciences 103: 9017-9022.

92. Dhanabal M, Wu F, Alvarez E, McQueeney KD, Jeffers M, et al. (2005) Recombinant semaphorin 6A-1 ectodomain inhibits in vivo growth factor and tumor cell line-induced angiogenesis. Cancer Biol Ther 4: 659-668.

93. Urbich C, Kaluza D, Fromel T, Knau A, Bennewitz K, et al. (2012) MicroRNA-27a/b controls endothelial cell repulsion and angiogenesis by targeting semaphorin 6A. Blood 119: 1607-1616.

94. Kigel B, Rabinowicz N, Varshavsky A, Kessler O, Neufeld G (2011) Plexin-A4 promotes tumor progression and tumor angiogenesis by enhancement of VEGF and bFGF signaling. Blood 118: 4285-4296.

95. Catalano A, Lazzarini R, Di Nuzzo S, Orciari S, Procopio A (2009) The plexin-A1 receptor activates vascular endothelial growth factor-receptor 2 and nuclear factor-kappaB to mediate survival and anchorage-independent growth of malignant mesothelioma cells. Cancer Res 69: 1485-1493.

96. Ghanem RC, Han KY, Rojas J, Ozturk O, Kim DJ, et al. (2011) Semaphorin 7A promotes angiogenesis in an experimental corneal neovascularization model. Curr Eye Res 36: 989-996.

97. Linderholm BK, Hellborg H, Johansson U, Elmberger G, Skoog L, et al. (2009) Significantly higher levels of vascular endothelial growth factor (VEGF) and shorter survival times for patients with primary operable triple-negative breast cancer. Ann Oncol 20: 1639-1646.

98. Ryden L, Jirstrom K, Haglund M, Stal O, Ferno M (2010) Epidermal growth factor receptor and vascular endothelial growth factor receptor 2 are specific biomarkers in triple-negative breast cancer. Results from a controlled randomized trial with long-term follow-up. Breast Cancer Res Treat 120: 491-498.

99. Liu HT, Ma R, Yang QF, Du G, Zhang CJ (2009) Lymphangiogenic characteristics of triple negativity in node-negative breast cancer. Int J Surg Pathol 17: 426-431.

100. Staton CA, Shaw LA, Valluru M, Hoh L, Koay I, et al. (2011) Expression of class 3 semaphorins and their receptors in human breast neoplasia. Histopathology 59: 274-282.

101. Gabrovska PN, Smith RA, Tiang T, Weinstein SR, Haupt LM, et al. (2011) Semaphorin-plexin signalling genes associated with human breast tumourigenesis. Gene 489: 63-69.

102. Castro-Rivera E, Ran S, Thorpe P, Minna JD (2004) Semaphorin 3B (SEMA3B) induces apoptosis in lung and breast cancer, whereas VEGF165 antagonizes this effect. Proc Natl Acad Sci U S A 101: 11432-11437.

103. Castro-Rivera E, Ran S, Brekken RA, Minna JD (2008) Semaphorin 3B inhibits the phosphatidylinositol 3-kinase/Akt pathway through neuropilin-1 in lung and breast cancer cells. Cancer Res 68: 8295-8303.

104. Miles DW, Chan A, Dirix LY, Cortes J, Pivot X, et al. (2010) Phase III study of bevacizumab plus docetaxel compared with placebo plus docetaxel for the first-line treatment of human epidermal growth factor receptor 2-negative metastatic breast cancer. J Clin Oncol 28: 3239-3247.

105. Bear HD, Tang G, Rastogi P, Geyer CE, Jr., Robidoux A, et al. (2012) Bevacizumab added to neoadjuvant chemotherapy for breast cancer. N Engl J Med 366: 310-320.

106. von Minckwitz G, Eidtmann H, Rezai M, Fasching PA, Tesch H, et al. (2012) Neoadjuvant chemotherapy and bevacizumab for HER2-negative breast cancer. N Engl J Med 366: 299-309.

107. Jubb AM, Miller KD, Rugo HS, Harris AL, Chen D, et al. (2011) Impact of exploratory biomarkers on the treatment effect of bevacizumab in metastatic breast cancer. Clin Cancer Res 17: 372-381.

108. Jubb AM, Strickland LA, Liu SD, Mak J, Schmidt M, et al. (2012) Neuropilin-1 expression in cancer and development. J Pathol 226: 50-60.

109. Jubb AM, Sa SM, Ratti N, Strickland LA, Schmidt M, et al. (2012) Neuropilin-2 expression in cancer. Histopathology.

110. Rody A, Holtrich U, Gaetje R, Gehrmann M, Engels K, et al. (2007) Poor Outcome in Estrogen Receptor-Positive Breast Cancers Predicted by Loss of Plexin B1. Clinical Cancer Research 13: 1115-1122.

111. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, et al. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10: 515-527.

112. Heiser LM, Sadanandam A, Kuo WL, Benz SC, Goldstein TC, et al. (2012) Subtype and pathway specific responses to anticancer compounds in breast cancer. Proc Natl Acad Sci U S A 109: 2724-2729.

113. Mac Gabhann F, Popel AS (2006) Targeting neuropilin-1 to inhibit VEGF signaling in cancer: Comparison of therapeutic approaches. PLoS Comput Biol 2: e180.

114. Stefanini MO, Wu FTH, Mac Gabhann F, Popel AS (2008) A compartment model of VEGF distribution in blood, healthy and diseased tissues. BMC Systems Biology 2: 77.

115. Segarra M, Ohnuki H, Maric D, Salvucci O, Hou X, et al. (2012) Semaphorin 6A regulates angiogenesis by modulating VEGF signaling. Blood 120: 4104-4115.

116. Ding Y, He D, Florentin D, Frolov A, Hilsenbeck S, et al. (2013) Semaphorin 4F as a critical regulator of neuroepithelial interactions and a biomarker of aggressive prostate cancer. Clin Cancer Res 19: 6101-6111.

117. Hochberg Y, Benjamini Y (1990) More powerful procedures for multiple significance testing. Stat Med 9: 811-818.

118. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, et al. (2010) Integrative genomic profiling of human prostate cancer. Cancer Cell 18: 11-22.

119. Yu YP, Landsittel D, Jing L, Nelson J, Ren B, et al. (2004) Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. J Clin Oncol 22: 2790-2799.

120. Cai C, Wang H, He HH, Chen S, He L, et al. (2013) ERG induces androgen receptor-mediated regulation of SOX9 in prostate cancer. J Clin Invest 123: 1109-1122.

121. Aryee MJ, Liu W, Engelmann JC, Nuhn P, Gurel M, et al. (2013) DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. Sci Transl Med 5: 169ra110.

122. Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, et al. (2012) The mutational landscape of lethal castration-resistant prostate cancer. Nature 487: 239-243.

123. Wu FT, Stefanini MO, Mac Gabhann F, Popel AS (2009) A compartment model of VEGF distribution in humans in the presence of soluble VEGF receptor-1 acting as a ligand trap. PLoS One 4: e5108.

124. Mac Gabhann F, Popel AS (2007) Interactions of VEGF isoforms with VEGFR-1, VEGFR-2, and neuropilin in vivo: a computational model of human skeletal muscle. Am J Physiol Heart Circ Physiol 292: H459-474.

125. Park JE, Chen HH, Winer J, Houck KA, Ferrara N (1994) Placenta growth factor. Potentiation of vascular endothelial growth factor bioactivity, in vitro and in vivo, and high affinity binding to Flt-1 but not to Flk-1/KDR. J Biol Chem 269: 25646-25654.

126. Joukov V, Sorsa T, Kumar V, Jeltsch M, Claesson-Welsh L, et al. (1997) Proteolytic processing regulates receptor specificity and activity of VEGF-C. EMBO J 16: 3898-3911.

127. Takahashi T, Fournier A, Nakamura F, Wang LH, Murakami Y, et al. (1999) Plexin-neuropilin-1 complexes form functional semaphorin-3A receptors. Cell 99: 59-69.

128. Martino MM, Briquez PS, Guc E, Tortelli F, Kilarski WW, et al. (2014) Growth factors engineered for super-affinity to the extracellular matrix enhance tissue healing. Science 343: 885-888.

129. Wu FT, Stefanini MO, Mac Gabhann F, Kontos CD, Annex BH, et al. (2010) A systems biology perspective on sVEGFR1: its biological function, pathogenic role and therapeutic use. J Cell Mol Med 14: 528-552.

130. Thielemann A, Baszczuk A, Kopczynski Z, Kopczynski P, Grodecka-Gazdecka S (2013) Clinical usefulness of assessing VEGF and soluble receptors sVEGFR-1 and sVEGFR-2 in women with breast cancer. Ann Agric Environ Med 20: 293-297.

131. Aoyagi Y, Iinuma H, Horiuchi A, Shimada R, Watanabe T (2010) Association of plasma VEGF-A, soluble VEGFR-1 and VEGFR-2 levels and clinical response and survival in advanced colorectal cancer patients receiving bevacizumab with modified FOLFOX6. Oncol Lett 1: 253-259.

132. Rini BI, Michaelson MD, Rosenberg JE, Bukowski RM, Sosman JA, et al. (2008) Antitumor activity and biomarker analysis of sunitinib in patients with bevacizumab-refractory metastatic renal cell carcinoma. J Clin Oncol 26: 3743-3748.

133. Willett CG, Boucher Y, Duda DG, di Tomaso E, Munn LL, et al. (2005) Surrogate markers for antiangiogenic therapy and dose-limiting toxicities for bevacizumab with radiation and chemotherapy: continued experience of a phase I trial in rectal cancer patients. J Clin Oncol 23: 8136-8139.

134. Li X (2010) VEGF-B: a thing of beauty. Cell Res 20: 741-744.

135. Kummel S, Eggemann H, Luftner D, Thomas A, Jeschke S, et al. (2006) Changes in the circulating plasma levels of VEGF and VEGF-D after adjuvant chemotherapy in patients with breast cancer and 1 to 3 positive lymph nodes. Anticancer Res 26: 1719-1726.

136. Vadasz Z, Haj T, Halasz K, Rosner I, Slobodin G, et al. (2012) Semaphorin 3A is a marker for disease activity and a potential immunoregulator in systemic lupus erythematosus. Arthritis Res Ther 14: R146.

137. Reuter CW, Morgan MA, Grunwald V, Herrmann TR, Burchardt M, et al. (2007) Targeting vascular endothelial growth factor (VEGF)-receptor-signaling in renal cell carcinoma. World J Urol 25: 59-72.

138. Lee E, Pandey NB, Popel AS (2013) Pre-treatment of mice with tumor-conditioned media accelerates metastasis to lymph nodes and lungs: a new spontaneous breast cancer metastasis model. Clin Exp Metastasis.

139. Mucci LA, Powolny A, Giovannucci E, Liao Z, Kenfield SA, et al. (2009) Prospective study of prostate tumor angiogenesis and cancer-specific mortality in the health professionals follow-up study. J Clin Oncol 27: 5627-5633.

140. Tretiakova M, Antic T, Binder D, Kocherginsky M, Liao C, et al. (2013) Microvessel density is not increased in prostate cancer: digital imaging of routine sections and tissue microarrays. Hum Pathol 44: 495-502.

141. Bono AV, Celato N, Cova V, Salvadore M, Chinetti S, et al. (2002) Microvessel density in prostate carcinoma. Prostate Cancer Prostatic Dis 5: 123-127.

142. Erbersdobler A, Isbarn H, Dix K, Steiner I, Schlomm T, et al. (2010) Prognostic value of microvessel density in prostate cancer: a tissue microarray study. World J Urol 28: 687-692.

143. Schoenborn JR, Nelson P, Fang M (2013) Genomic profiling defines subtypes of prostate cancer with the potential for therapeutic stratification. Clin Cancer Res 19: 4058-4066.

144. Duque JL, Loughlin KR, Adam RM, Kantoff PW, Zurakowski D, et al. (1999) Plasma levels of vascular endothelial growth factor are increased in patients with metastatic prostate cancer. Urology 54: 523-527.

145. Caine GJ, Blann AD, Stonelake PS, Ryan P, Lip GY (2003) Plasma angiopoietin-1, angiopoietin-2 and Tie-2 in breast and prostate cancer: a comparison with VEGF and Flt-1. Eur J Clin Invest 33: 883-890.

146. Singh A, Gautam KA, Dalela D, Sankhwar S, Natu S, et al. (2013) Plasma vascular endothelial growth factors A and C in patients undergoing prostatic biopsy and TURP for suspected prostatic neoplasia. Asian Pac J Cancer Prev 14: 2053-2058.

147. Saylor PJ, Kozak KR, Smith MR, Ancukiewicz MA, Efstathiou JA, et al. (2012) Changes in biomarkers of inflammation and angiogenesis during androgen deprivation therapy for prostate cancer. Oncologist 17: 212-219.

148. Fontana A, Galli L, Fioravanti A, Orlandi P, Galli C, et al. (2009) Clinical and pharmacodynamic evaluation of metronomic cyclophosphamide, celecoxib, and dexamethasone in advanced hormone-refractory prostate cancer. Clin Cancer Res 15: 4954-4962.

149. Tomic TT, Gustavsson H, Wang W, Jennbacken K, Welen K, et al. (2012) Castration resistant prostate cancer is associated with increased blood vessel stabilization and elevated levels of VEGF and Ang-2. Prostate 72: 705-712.

150. Tannock IF, Fizazi K, Ivanov S, Karlsson CT, Flechon A, et al. (2013) Aflibercept versus placebo in combination with docetaxel and prednisone for treatment of men with metastatic castration-resistant prostate cancer (VENICE): a phase 3, double-blind randomised trial. Lancet Oncol 14: 760-768.

151. Beardsley EK, Hotte SJ, North S, Ellard SL, Winquist E, et al. (2012) A phase II study of sorafenib in combination with bicalutamide in patients with chemotherapy-naive castration resistant prostate cancer. Invest New Drugs 30: 1652-1659.

152. Michaelson MD, Oudard S, Ou YC, Sengelov L, Saad F, et al. (2014) Randomized, Placebo-Controlled, Phase III Trial of Sunitinib Plus Prednisone Versus Prednisone Alone in Progressive, Metastatic, Castration-Resistant Prostate Cancer. J Clin Oncol 32: 76-82.

153. Smith DC, Smith MR, Sweeney C, Elfiky AA, Logothetis C, et al. (2013) Cabozantinib in patients with advanced prostate cancer: results of a phase II randomized discontinuation trial. J Clin Oncol 31: 412-419.

154. Bismar TA, Demichelis F, Riva A, Kim R, Varambally S, et al. (2006) Defining aggressive prostate cancer using a 12-gene model. Neoplasia 8: 59-68.

155. Erho N, Crisan A, Vergara IA, Mitra AP, Ghadessi M, et al. (2013) Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. PLoS One 8: e66855.

156. Irshad S, Bansal M, Castillo-Martin M, Zheng T, Aytes A, et al. (2013) A molecular signature predictive of indolent prostate cancer. Sci Transl Med 5: 202ra122.

157. Burton JB, Priceman SJ, Sung JL, Brakenhielm E, An DS, et al. (2008) Suppression of prostate cancer nodal and systemic metastasis by blockade of the lymphangiogenic axis. Cancer Res 68: 7828-7837.

158. Jennbacken K, Vallbo C, Wang W, Damber JE (2005) Expression of vascular endothelial growth factor C (VEGF-C) and VEGF receptor-3 in human prostate cancer is associated with regional lymph node metastasis. Prostate 65: 110-116.

159. Aihara M, Wheeler TM, Ohori M, Scardino PT (1994) Heterogeneity of prostate cancer in radical prostatectomy specimens. Urology 43: 60-66; discussion 66-67.

160. Liu W, Laitinen S, Khan S, Vihinen M, Kowalski J, et al. (2009) Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer. Nat Med 15: 559-565.

161. Anisimov A, Leppanen VM, Tvorogov D, Zarkada G, Jeltsch M, et al. (2013) The basis for the distinct biological activities of vascular endothelial growth factor receptor-1 ligands. Sci Signal 6: ra52.

162. Patnaik A, LoRusso PM, Messersmith WA, Papadopoulos KP, Gore L, et al. (2014) A Phase Ib study evaluating MNRP1685A, a fully human anti-NRP1 monoclonal antibody, in combination with bevacizumab and paclitaxel in patients with advanced solid tumors. Cancer Chemother Pharmacol 73: 951-960.

163. Cunningham D, Humblet Y, Siena S, Khayat D, Bleiberg H, et al. (2004) Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. N Engl J Med 351: 337-345.

164. Schoeberl B, Faber AC, Li D, Liang MC, Crosby K, et al. (2010) An ErbB3 antibody, MM-121, is active in cancers with ligand-dependent activation. Cancer Res 70: 2485-2494.

165. Shepherd FA, Rodrigues Pereira J, Ciuleanu T, Tan EH, Hirsh V, et al. (2005) Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med 353: 123-132.

166. Cameron D, Casey M, Press M, Lindquist D, Pienkowski T, et al. (2008) A phase III randomized comparison of lapatinib plus capecitabine versus capecitabine alone in women with advanced breast cancer that has progressed on trastuzumab: updated efficacy and biomarker analyses. Breast Cancer Res Treat 112: 533-543.

167. Arnould L, Arveux P, Couturier J, Gelly-Marty M, Loustalot C, et al. (2007) Pathologic complete response to trastuzumab-based neoadjuvant therapy is related to the level of HER-2 amplification. Clin Cancer Res 13: 6404-6409.

168. Perez EA, Romond EH, Suman VJ, Jeong JH, Davidson NE, et al. (2011) Four-year follow-up of trastuzumab plus adjuvant chemotherapy for operable human epidermal growth factor receptor 2-positive breast cancer: joint analysis of data from NCCTG N9831 and NSABP B-31. J Clin Oncol 29: 3366-3373.

169. Khambata-Ford S, Garrett CR, Meropol NJ, Basik M, Harbison CT, et al. (2007) Expression of epiregulin and amphiregulin and K-ras mutation status predict disease control in metastatic colorectal cancer patients treated with cetuximab. J Clin Oncol 25: 3230-3237.

170. Yonesaka K, Zejnullahu K, Okamoto I, Satoh T, Cappuzzo F, et al. (2011) Activation of ERBB2 signaling causes resistance to the EGFR-directed therapeutic antibody cetuximab. Sci Transl Med 3: 99ra86.

171. Ghosh R, Narasanna A, Wang SE, Liu S, Chakrabarty A, et al. (2011) Trastuzumab has preferential activity against breast cancers driven by HER2 homodimers. Cancer Res 71: 1871-1882.

172. Shankaran H, Zhang Y, Tan Y, Resat H (2013) Model-based analysis of HER activation in cells co-expressing EGFR, HER2 and HER3. PLoS Comput Biol 9: e1003201.

173. Schoeberl B, Pace EA, Fitzgerald JB, Harms BD, Xu L, et al. (2009) Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis. Sci Signal 2: ra31.

174. Hudis CA (2007) Trastuzumab--mechanism of action and use in clinical practice. N Engl J Med 357: 39-51.

175. Press WH (1992) Numerical recipes in FORTRAN : the art of scientific computing. Cambridge [England] ; New York, NY, USA: Cambridge University Press. xxvi, 963 p. p.

176. Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nat Biotechnol 20: 370-375.

177. Chen WW, Schoeberl B, Jasper PJ, Niepel M, Nielsen UB, et al. (2009) Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. Mol Syst Biol 5: 239.

178. Roepstorff K, Grandal MV, Henriksen L, Knudsen SL, Lerdrup M, et al. (2009) Differential effects of EGFR ligands on endocytic sorting of the receptor. Traffic 10: 1115-1127.

179. Wang Y, Tian T, Hu Z, Tang J, Wang S, et al. (2008) EGF promoter SNPs, plasma EGF levels and risk of breast cancer in Chinese women. Breast Cancer Res Treat 111: 321-327.

180. Chien CH, Huang CC, Lin YH, Shen J, Chow SN (1997) Detection of serum transforming growth factor-alpha in patients of primary epithelial ovarian cancers by enzyme immunoassay. Gynecol Oncol 66: 405-410.

181. Ritter CA, Perez-Torres M, Rinehart C, Guix M, Dugger T, et al. (2007) Human breast cancer cells selected for resistance to trastuzumab in vivo overexpress epidermal growth factor receptor and ErbB ligands and remain dependent on the ErbB receptor network. Clin Cancer Res 13: 4909-4919.

182. Niepel M, Hafner M, Pace EA, Chung M, Chai DH, et al. (2013) Profiles of Basal and stimulated receptor signaling networks predict drug response in breast cancer lines. Sci Signal 6: ra84.

Robert Joseph Bender
bender.rj@gmail.com

EDUCATION

2014 **Ph.D. in Biomedical Engineering**
The Johns Hopkins University, Baltimore, MD
Thesis: "Genomic Data-Based Models of Growth Factor Signaling for Personalized Cancer Therapy Selection"

2010 **M.S. in Biotechnology**
The Johns Hopkins University, Baltimore, MD
Thesis: "Effects of Amyloid Beta Fibrils and Aggregation Intermediates on Giant Unilamellar Vesicles"

2006 **B.S. in Chemical Engineering**
University of Maryland, College Park, MD
Cum Laude Graduate
Certificate in Biological Engineering

RESEARCH EXPERIENCE

2010-present Ph.D. Candidate, The Johns Hopkins University
Department of Biomedical Engineering and Institute for Computational Medicine
Advisor: Dr. Feilim Mac Gabhann
Research: Computational modeling of growth factor signaling pathways related to angiogenesis. Use of patient-specific gene expression data to predict growth factor pathway activity and drug response in tumors. Integrated analysis of RNA-Seq, microRNA-Seq, and reverse phase protein array data to predict effects of microRNA on protein levels. *In vitro* validation of computational models in endothelial cell lines.

2009-2010 Masters Student, The Johns Hopkins University
Thesis research conducted in the Fischell Department of Bioengineering, University of Maryland
Advisor: Dr. Sameer Shah
Research: Synthesized giant unilamellar vesicles and studied the effects of adverse osmotic environments on the mechanical properties of their membranes. Utilized giant unilamellar vesicles as model cell membranes to determine the impact that amyloid beta-peptide monomers, oligomers, and fibrils have in Alzheimer's disease. Developed a method for detecting the membranes of vesicles using the Matlab Image Processing Toolbox.

2005-2006 Undergraduate Researcher, University of Maryland
Department of Chemical Engineering
Advisor: Dr. Srinivasa Raghavan
Research: Prepared various self-assembled structures, including lipid/fatty acid vesicles, nanocapsules, and gels. Performed experiments to determine the release kinetics of small molecules from self-assembled systems. Studied the applicability of these systems for drug delivery and controlled release.

PROFESSIONAL EXPERIENCE

2006-2010 Bioprocess Engineer, Human Genome Sciences, Inc, Rockville, MD
Department of Fermentation and Cell Culture Sciences
Responsibilities: Developed microbial and mammalian cell culture processes for production of recombinant proteins. Performed process characterization and validation studies to support regulatory filings. Statistically designed experiments, using approaches that included full and partial factorial design and response surface methodology. Developed harvest and recovery processes, including continuous centrifugation, depth filtration, and tangential flow filtration. Participated in technical transfers of cell culture and recovery processes to manufacturing facilities. Provided on-site support to GMP manufacturing campaigns. Modeled mixing conditions in bioreactors using computational fluid dynamics. Attended professional conferences.

PUBLICATIONS

Bender RJ and Mac Gabhann F. "Personalized computational models of EGFR/ErbB family signaling reveal mechanisms of intrinsic resistance in patient subgroups." In preparation.

Bender RJ and Mac Gabhann F. "Dysregulation of the vascular endothelial growth factor and semaphorin ligand-receptor families in prostate cancer metastasis." Under review.

Bender RJ and Mac Gabhann F. "Expression of VEGF and semaphorin genes defines subgroups of triple-negative breast cancer." *PLoS One* 8(5): e61788 (2013).

PRESENTATIONS

Hosseini I, **Bender RJ**, and Mac Gabhann F. "Identification of gene biomarkers for anti-cancer drugs using parallel PLS-based models." Poster presented at the 18th Annual International Conference on Research in Computational Molecular Biology, Pittsburgh, PA, April 3, 2014.

Bender RJ and Mac Gabhann F. "VEGF and semaphorin expression in breast cancer: harnessing high-throughput data for improved targeting of angiogenesis." Presentation at the Biomedical Engineering Society (BMES) 2012 Annual Meeting, Atlanta, GA, October 26, 2012.

Bender RJ and Mac Gabhann F. "VEGF and semaphorin expression in breast cancer: defining anti-angiogenesis treatment subgroups." Poster presented at the Johns Hopkins 6th Annual Nano-Bio Symposium, Baltimore, MD, May 4, 2012.

TEACHING EXPERIENCE

Fall 2012 TA for EN.580.429 Systems Bioengineering III

Spring 2013 TA for EN.580.200 Introduction to Scientific Computing in BME using Python, Matlab, and R